Modeling and Understanding Aqueous Mixtures Using Kirkwood-Buff Theory of Solutions

Vom Fachbereich Chemie
der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines
Doktor rerum naturalium (Dr. rer. nat.)

genehmigte
Dissertation

vorgelegt von

Pritam Ganguly,
Master of Science (M. Sc.)

aus Makardaha, Howrah, Indien

Referent: Prof. Dr. Nico F. A. van der Vegt
Korreferent: Prof. Dr. Robert Berger
Tag der Einreichung: 30. Oktober 2013
Tag der mündlichen Prüfung: 12. Dezember 2013

Darmstadt, 2014
D17

to

my parents

and grandparents

Contents

Summary v

Zusammenfassung ix

Acknowledgments xiii

Publications xv

1 Introduction 1

2 Kirkwood-Buff Theory of Solutions 7
- 2.1 Kirkwood-Buff Integrals 8
- 2.2 KBIs and Thermodynamic Quantities 8
  - 2.2.1 Binary mixtures 9
  - 2.2.2 Ternary mixtures 10
  - 2.2.3 Inversion of the Kirkwood-Buff theory 14

3 Kirkwood-Buff Theory and Computer Simulations of Aqueous Mixtures -a Review 17
- 3.1 Introduction 18
- 3.2 Kirkwood-Buff Derived Force-fields for Aqueous Solutions 20
- 3.3 Kirkwood-Buff Theory and Solvation Thermodynamics from Computer Simulations 23
- 3.4 Computation of Kirkwood-Buff Integrals and Technical Issues 25
- 3.5 Conclusions 28

4 Kirkwood-Buff Coarse-grained Force-fields for Aqueous Solutions 37
- 4.1 Introduction 38
- 4.2 Computational Details 40
- 4.3 Results and Discussions 41
  - 4.3.1 Method and Implementation 41

4.3.2 Applications . . . 42
4.4 Conclusions . . . 49
4.5 Supporting Information of Chapter 4 . . . 50
4.5.1 Pre-factors "A" used in the ramp potential . . . 50
4.5.1.1 Urea-Water (parameterization at 4.7 M) . . . 50
4.5.1.2 Benzene-Water (parameterization at 0.2 M) . . . 51
4.5.2 Urea-Water simulated with MARTINI model . . . 51
4.5.3 KB-IBI potentials for urea-water at different urea concentrations 52
4.6 Acknowledgements . . . 53

5 Representability and Transferability of KB-IBI Models for Multicomponent Aqueous Systems 57
5.1 Introduction . . . 59
5.2 Kirkwood-Buff Iterative Boltzmann Inversion . . . 61
5.3 Computational Details . . . 62
5.4 Results and Discussion . . . 63
5.4.1 Salting-in at infinite solute dilution . . . 63
5.4.2 Salting-in of benzene at finite concentrations . . . 68
5.4.3 Cluster analysis of benzene in urea-water solution . . . 75
5.4.4 Structure-based coarse-graining with additional thermodynamic targets . . . 79
5.4.4.1 Pure water . . . 79
5.4.5 Binary urea-water mixture . . . 80
5.4.6 Benzene in urea-water at infinite dilution of benzene . . . 84
5.5 Conclusions . . . 86

6 Ion Pairing in Aqueous Electrolyte Solutions with Biologically Relevant Anions 93
6.1 Introduction . . . 94
6.2 Computational Details . . . 96
6.3 Results and Discussions . . . 98
6.4 Conclusions . . . 104
6.5 Acknowledgment . . . 105

7 Salting-out of Benzene with Hofmeister Cations 109
7.1 Introduction . . . 111

7.2 Thermodynamic Theory 112
7.3 Computational Details 114
7.4 Results and Discussion 116
7.5 Discussion and Conclusion 127
7.6 Acknowledgement 129

8 Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions 133
8.1 Introduction 134
8.2 Computational Details 137
8.3 Results and Discussions 137
8.3.1 Convergence of KBIs with simulation time 137
8.3.2 Urea-water KBIs from large system sizes 139
8.3.3 Methanol-water KBIs from large system sizes 144
8.3.4 KBIs with varying simulation box size 145
8.3.4.1 Effect of RDF tail corrections on RKBIs of small and large methanol-water systems 145
8.3.4.2 Effect of RDF tail corrections on RKBIs of small and large urea-water systems 149
8.4 Conclusions 151
8.5 Acknowledgements 152

9 Conclusions and Outlook 155

Summary

Fluctuation theory of solutions, introduced by Kirkwood and Buff in 1951, relates particle number fluctuations of small scales to the global thermodynamic properties of the system. Cosolvent effects on (bio)solutes in aqueous solutions can be modeled in molecular dynamics simulations and the fluctuation theory of solutions, termed as Kirkwood-Buff (KB) theory, provides a theoretical route to analyze the specific interactions between the solute and the cosolvent in terms of the preferential solvation or preferential binding which is important to study, amongst others, the conformational changes in biomacromolecules under different cosolvent conditions. The KB theory can be used to relate the simulation data on the integrals of the pair-correlation functions between the system components over volume to the experimentally observable thermodynamic quantities such as change in the chemical potential of the solutes with varying cosolvent concentrations. Although computer simulations provide the interaction mechanisms or the dynamics of a system at molecular level, but they are confronted with many challenges related to the limitations in computational power, accuracy of the models and straight-forward comparison with experimental data. The number of particles in a system can be reduced significantly by grouping several particles in single interaction-sites, termed as coarse-grained beads, which leads to significant speed-up of the simulations. With coarse-graining, simulations of larger systems with longer time-scale are possible which are required for the most of the biological processes. In this thesis we use the KB theory to develop simplified coarse-grained models for aqueous binary and ternary mixtures. On the other hand, by resolving the integrals of the pair-correlation functions to the contributions arising from different modes of spatial separation between the solution components, the KB theory is also used to explain the ion-specific pairing mechanisms between Hofmeister ions and the ion-specific changes in the solvation thermodynamics of solutes in aqueous solutions with all-atom simulations.

This thesis includes a theoretical account on the Kirkwood-Buff theory explaining the relevant thermodynamic relations which is followed by a review on the applications of the Kirkwood-Buff theory to the computer simulations of aqueous solutions. Then this thesis proposes a new method of coarse-graining by combining

the structure-based Iterative Boltzmann Inversion (IBI) method and the Kirkwood-Buff (KB) theory. The method, KB-IBI, is applied to binary mixtures of urea-water and benzene-water and the single-site coarse-grained potentials for the molecules are found to be consistent with the atomistic pair-correlations and the variations in the urea or benzene chemical potentials with different solution concentrations. As urea serves as a chemical denaturant for proteins, application of these coarse-grained potentials to the ternary mixtures of solutes in urea-water would be the first step towards modeling the urea-driven conformational changes in biomolecules. So the preferential interactions between benzene and urea are studied with single-site coarse-grained models and the variation in the solvation free-energy of benzene with different urea concentrations has been reproduced in agreement with the atomistic model. The representability and the convergence of the KB-IBI coarse-grained models at a particular state-point where the model is parametrized are discussed in terms of the thermodynamic quantities such as pressure, potential energy and the variation in the solvation free-energy for the systems of pure water, binary urea-water mixture and ternary benzene-urea-water mixtures at infinite benzene dilution. The transferability issue of the KB-IBI potentials at different urea concentrations has also been examined and a cluster analysis of benzene in urea-water solutions is discussed.

With all-atomistic simulations the application of the KB theory in an analysis of monovalent alkali cation pairing with biologically relevant anions such as acetate or phosphate has revealed a ion-specific variation in the water-mediated ion-pairs which leads to the variation in the activity of the salts. Contributions to the integrals of the pair-correlation functions originating from the different ion pairing modes, namely contact ion-pairs (direct pairing between the cation and the anion), solvent-shared ion-pairs (solvation-shells of the ions are shared) or solvent-separated ion-pairs (solvation-shells of the ions are separated), have been analysed. It has been found that solvent-separated ion-pairing mechanism for phosphate and solvent-shared mechanism for acetate play the major role in the ion-specific changes in the salt activity in the solution; whereas for chloride solutions contact ion-pairing mechanism prevails over solvent-mediated mechanisms. For the ternary systems of solutes in the salt-solutions, the interactions between benzene and the ions in aqueous solutions of the alkali chlorides have been studied with KB theory and different force-fields models have been tested. Simulation data suggest that the direct correlations between benzene and ions play more significant role rather than the indirect ion-pairing to explain the ion-specific decrease in the solubility of benzene, termed

as ion-specific salting-out of benzene, upon addition of salts. A geometric packing of hydrated lithium ions around benzene is found to be the reason of lithium chloride being less salting-out agent than sodium chloride or potassium chloride. Calculation of the integrals of the pair-correlation functions over volume, termed as Kirkwood-Buff integrals (KBIs) and which are the key quantities in the KB theory to relate the local pair-structures to the thermodynamic quantities, does come with many technical issues. Calculation of more precise KBIs and the effect of the system size and the simulation time on the KBIs are discussed with the help of binary mixtures of urea-water and methanol-water where the convergence issues of the KBIs are more pronounced due to the microheterogeneity of the solutions and slower dynamics of the local domains.

This thesis serves as an account on the diverse applicability of KB theory to the computer simulations of biologically important systems. The newly developed coarse-graining method, KB-IBI, can be thought as a novel step towards modeling aqueous single-phase solutions and can potentially be extended to model polymers or biomacromolecules in urea-water or other cosolvent-water solutions. Also the KB theory in general can be used to quantify the preferential solvation of the solutes with cosolvents and to study the conformational changes in the biomolecules, provided that the technical issues in the calculation of the KBIs are addressed properly.

Zusammenfassung

Die Fluktuationstheorie von Lösungen verknüpft Teilchenfluktuationen und Paarkorrelationen zwischen den Partikel auf einer lokalen Ebene mit den globalen thermodynamischen Eigenschaften. Co-Lösungsmittel, welche das Verhalten von gelösten (Bio)Molekülen in wässriger Lösung beeinflussen, können in molekulardynamischen Simulationen modelliert werden. Die Fluktuationstheorie von Lösungen ermöglicht eine theoretische Beschreibung um die spezifischen Wechselwirkungen zwischen gelöstem Teilchen und Co-Lösungsmittel zu beschreiben. Man erhält so Informationen über bevorzugte Solvatation oder bevorzugte Bindungsbildung. Dies ist sehr wichtig um Konformationsänderungen in Biomakromolekülen unter verschiedenen Co-Lösungsmittelbedingungen zu untersuchen. Obwohl mit Hilfe von Computersimulationen Wechselwirkungsmechanismen oder die Dynamik eines Systems auf einer molekularen Ebene berechnet werden können, gibt es aufgrund gewisser Einschränkungen, wie z.B. des hohen rechnerischen Aufwandes, der Genauigkeit der verwendeten Modelle oder des Vergleichs mit experimentellen Daten, Herausforderungen, die gelöst werden müssen. Die Anzahl an Teilchen in einem System lässt sich signifikant verringern, indem man mehrere Atome zu einem Superatom zusammenfasst, eine sog. Vergröberungskugel. Dies führt zu einer erheblichen Beschleunigung der Simulationen. Mit dieser Vergröberung des Systems lassen sich größere Systeme auf einer längeren Zeitskala simulieren, welche zur Untersuchung der meisten biologischen Prozesse notwendig ist. Die Fluktuationstheorie von Lösungen, erstmals vorgestellt von Kirkwood und Buff im Jahre 1951, kann dazu genutzt werden, um die aus den Simulationen erhaltenen Daten über Paarstrukturen der Systeme mit experimentell zugänglichen thermodynamischen Größen zu verknüpfen. Dazu gehört die änderung des chemischen Potentials des gelösten Teilchens bei unterschiedlichen Konzentrationen des Co-Lösungsmittels. Ziel dieser Arbeit ist es, mit Hilfe der Kirkwood-Buff Theorie Vergröberungsmodelle für binäre und ternäre wässrige Mischungen zu entwickeln. Darüber hinaus dient sie zur Erklärung der Mechanismen der Ionenpaarung und der Wechselwirkungen der gelösten Stoffe mit den Ionen auf einer atomistischen Skala.

Die hier vorgelegte Arbeit liefert zunächst eine Zusammenfassung der Kirkwood-Buff Theorie um die relevanten thermodynamischen Beziehungen zu erklären. Des weiteren wird eine Nachbetrachtung der Anwendung der Kirkwood-Buff Theorie in Computersimulationen wässriger Lösungen präsentiert. Anschließend wird ein neues Vergröberungsmodell vorgeschlagen, welches die Struktur basierte Iterative Boltzmann Inversion (IBI) mit der Kirkwood-Buff (KB) Theorie verknüpft. Diese Methode, KB-IBI, wird auf binäre Mischungen aus Harnstoff-Wasser und Benzol-Wasser angewandt. Die Superatom-Vergröberungs-Potentiale der Moleküle stimmen mit den atomistischen Paarstrukturen und den Veränderungen des chemischen Potentials von Harnstoff oder Benzol, bei unterschiedlichen Konzentrationen der Lösungen, überein. Da Harnstoff als chemisches Denaturierungsmittel für Proteine dient, wäre ein erster Schritt um die von Harnstoff verursachten Konformationsänderungen in den Biomolekülen zu modellieren, die Anwendung dieser Vergröberungspotentiale auf Simulationen ternärer Systeme von gelösten Molekülen in Harnstoff-Wasser-Lösungen. Die bevorzugten Wechselwirkungen zwischen Benzol und Harnstoff werden also mit den Superatom-Vergröberungsmodellen untersucht. Die korrekte Veränderung der freien Lösungsenergie von Benzol, je nach Harnstoffkonzentration, konnte erfolgreich reproduziert werden. Die Darstellbarkeit und Konvergenz der KB-IBI-Vergröberungsmodelle an dem Zustandspunkt, welcher zur Parametrisierung des Models genutzt wurde, wird an Hand thermodynamischer Größen wie Druck, potentielle Energie oder der Veränderung der freien Lösungsenergie von reinem Wasser, binären Harnstoff- Wassermischungen und ternären Benzol-Harnstoff-Wassermischungen bei unendlicher Verdünnung von Benzol, diskutiert. Die übertragbarkeit der KB-IBI Potentiale bei verschiedenen Harnstoffkonzentrationen wurde ebenfalls im Detail untersucht. Auch wird eine Clusteranalyse von Benzol in Harnstoff-Wasser Lösungen diskutiert.

Rein atomistische Simulationen, um den Mechanismus der Ionenpaarung monovalenter Alkalikationen mit biologisch relevanten Anionen, wie Acetat oder Phosphat mittels KB-Theorie zu verstehen, haben gezeigt, dass in Wasser vorliegende Ionenpaare ionenspezifisch verändert werden, was zu einer änderung der Aktivität der Salze führt. Für ternäre Systeme aus gelösten Stoffen in Salzlösungen sind die Wechselwirkungen zwischen Benzol und den Ionen in wässriger Alkalichloridlösung mit der KB Theorie und verschiedenen Kraftfeldmodellen untersucht worden. Die erhaltenen Daten lassen darauf schließen, dass direkte Korrelationen zwischen Benzol und den Ionen eine größere Rolle spielen als die indirekte Ionenpaarung, um

nach Hinzufügen der Salze die ionenspezifische Erniedrigung der Löslichkeit von Benzol, das sog. ionenspezifische Aussalzen von Benzol ,zu erklären. Eine räumliche Ansammlung hydratisierter Lihium-Ionen um das Benzol ist als Grund für die geringere Fähigkeit des Aussalzens von Lithiumchlorid im Vergleich zu Natriumchlorid oder Kaliumchlorid ermittelt worden. Berechnungen der Integrale der Paarkorrelationsfunktionen über das Volumen, auch Kirkwood-Buff Integrale (KBI) genannt, stellen einige technische Anforderungen. Die KBI sind die Schlüsselgrößen in der KB Theorie, die die Zuordnung der lokalen Paarstrukturen zu den thermodynamischen Größen ermöglichen. Berechnungen genauerer KBI und der Einfluss der Größe des Systems, sowie der Simulationszeit werden an Hand von Simulationen binärer Mischungen von Harnstoff- Wasser und Methanol-Wasser diskutiert. In diesen Systemen sind Konvergenzprobleme besser zu ermitteln, da die Lösungen eine Mikroheterogenität aufweisen und die lokalen Einheiten einer langsameren Dynamik folgen.

Die hier vorliegende Arbeit dient als Zusammenstellung für die verschiedene Anwendbarkeit der KB Theorie für Computersimulationen biologisch wichtiger Systeme. Die neu entwickelte Vergröberungsmetode, KB-IBI, kann als ein neuer Schritt auf dem Weg einphasige, wässrige Lösungen zu modellieren und möglicherweise darüber hinaus zur Beschreibung von Polymeren oder Biomakromolekülen in Harnstoff- Wasser- oder anderen Co-Lösungsmittel-Wasser-Lösungen, gesehen werden. Allgemein kann die KB Theorie auch dazu genutzt werden um das bevorzugte Lösungsverhalten von gelösten Teilchen mit Co-Lösungsmitteln akkurat zu quantifizieren und Konformationsänderungen von Biomolekülen zu untersuchen, vorausgesetzt die technischen Probleme bei der Berechnung der KBI werden ansprechend gelöst.

Sincere Thanks to...

The journey was made out of moments; the moments and the people with whom I witnessed those moments deserve respect and gratitude.

I would like to convey my deepest respect and thanks to Prof. Dr. Nico van der Vegt for granting me the opportunity to pursue my Ph.D. studies in his group with his sincere guidance. During my Ph.D. years he has provided me with the fortuity to learn about the subject as much as I could. With that, his friendly gestures and academically supportive nature have always been very motivational! I would like to thank Prof. Dr. Florian Müller-Plathe and his group for providing me with the knowledge about the research going on in their group during the course of the theory-seminars. I would like to thank Prof. Dr. Robert Berger, apl. Prof. Michael C. Böhm, and PD Dr. Elmar Bonaccurso for being in my Ph.D. committee. My respect goes to Prof. Dr. Padma Kumar Padmanabhan for introducing me to the field of molecular dynamics simulations and for valuable discussions during my Ph.D. years. I thank all my teachers who shared their knowledge with me to enrich my thoughts.

Thanks go to my dear colleagues, ex-colleagues and collaborators. My Ph.D.-life would not have been the same without your constant involvement and support. Francisco, Pim, Timir, Debashish da, Emiliano, Fereshte, David, Gregor, Tingting, Valentina, Chunli, Pranay, Claudia, Samira, Imke, Bin, Ran and all other colleagues – all I can say is a big thanks! Thanks go to my friends in Darmsatdt – Upendra, Mani, Titu, Prashant, Aniruddh, Abhinav, Murthy, Rajesh, Sudheer, Ravi, Sandeep, Bro, Shawn, Evangelos, Henner, Abbas and all others. Thanks go to all of my school and college friends: Amitava, Bhaipo, Dipayan, Souvik, Suvomay, Gargi, Debjit, Biswa, Sourish, Abhik, Sagnik, Sukamal, Amit – I wish I could go on with the names! All my friends from IITG deserve a special thanks; the experience at IITG would not have been an experience at all without you: Pokkhi, Ahom, Sisir, Supriya da, Kausik da, Saty, Arghya, Tuhin, Chiki, Rajesh, Suresh, Guchhi, Jana, Gopi, Debu, Shyama, Madhu, Mankho, Rajorshi, Gora, Anka, Baswanth, Anup, Asim, Srijit, Sudip Dey, Sudip Sarkar, Pabitra, Prakash, Kapil Ji, Arun, Jithin, Deepak, Chandan, Konark, Pankaj, Piyali, Bhawana, Sukhi bhaiyya, Arnab da , Sabya da, Bedo da, Mrinmoy da,

Atreyi di, Ponchokesh, Jit, Tanmoy, Kartik, Sumit, Hridesh, Koushik, all my seniors, batch-mates and juniors and all the staffs of Siang hostel.

Oh, I was missing some names: Maa, Baba, Jethu, Amma, Bachchha, Naru, Bhoott-Darling, Polton, Ghaora, Gaurango, Imon di...

Pritam

Publications

This thesis is based on the following publications:

P. Ganguly, D. Mukherji, C. Junghans, and N. F. A. van der Vegt, *Kirkwood-Buff Coarse-Grained Force Fields for Aqueous Solutions*, J. Chem. Theory Comput., 8:1802 (2012).

P. Ganguly, and N. F. A. van der Vegt, *Representability and Transferability of Kirkwood-Buff Iterative Boltzmann Inversion Models for Multicomponent Aqueous Systems*, J. Chem. Theory Comput., 9:5247 (2013).

P. Ganguly, P. Schravendijk, B. Hess, and N. F. A. van der Vegt, *Ion Pairing in Aqueous Electrolyte Solutions with Biologically Relevant Anions*, J. Phys. Chem. B, 115:3734 (2011).

P. Ganguly, T. Hajari, and N. F. A. van der Vegt, *Salting-out of Benzene with Hofmeister Cations: an Insight to the Hydrophobic Ring-Cation Interactions*, (submitted).

P. Ganguly, and N. F. A. van der Vegt, *Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations*, J. Chem. Theory Comput., 9:1347 (2013).

1 Introduction

Aqueous solutions of biomolecules, or, more general, water soluble macromolecules show quite dramatic variation in their thermodynamic properties upon changing internal and external conditions such as temperature, pressure or other mechanical disturbances along with the changes in the chemical environment the systems have been subjected to. Understanding biological systems, which are mostly complex solutions with water being the major component, and their dependency on the chemical compositions of the solutions are often challenging. Experiments in laboratories provide interesting results regarding the stability of biological macromolecules, such as proteins, in different chemical environments, such as in different compositions and/or concentrations of salts in aqueous solutions,[1–4] but often fail to propose strong arguments regarding the underlying mechanisms behind these processes at molecular level. Computer simulations of the biological systems can provide the mechanics of the molecules and the physical explanations regarding these biological processes but often confront with many technical challenges. Limited computational resources do not allow simulation studies of the systems/processes with larger time and/or length scales. Low-resolution or coarser description of a system (coarse-graining) where groups of particles or atoms are treated as single interaction sites can help to overcome the challenges regarding time and length-scales by reducing the number of interactions in a system and by smoothening the free-energy landscape of the system so that the system evolves much faster.[5] Cosolvent-induced conformational changes in (bio)macromolecules require larger time and length-scales, so computational studies of these processes require simplified, coarse-grained models that retain chemical specificity. The coarse-grained models should mimic the changes in the molecular structures of the solutions and thermodynamic quantities related to the conformational changes of the solute macromolecules with varying cosolvent compositions in order to physically explain the effects of the cosolvents. But available coarse-graining methods often fail to model the cosolvent-effects in biological aqueous systems.

On the other hand, validation of the results obtained from computer simulations by comparing directly with experimental data is not always straightforward. A fluc-

tuation theory of solutions, which relates molecular distribution functions to thermodynamic quantities relevant to solvation and stability of solutes in solutions such as partial molar volumes or the variation of the chemical potential of the solutes with varying solution compositions, provides a route to compare the results obtained from computer simulations with experimental results in a computationally cheaper way. The theory, introduced by Kirkwood and Buff,[6] uses the integrals of the pair-correlation functions over volume, named as Kirkwood-Buff integrals, and links them to the thermodynamic quantities which are determined by the fluctuations of the statistical ensemble. The theory is exact and does not assume any pair-wise additivity of the interaction potentials of the system. An inverse theory of the Kirkwood-Buff theory, derived by Ben-Naim,[7] provides the equations to calculate the Kirkwood-Buff integrals from experimental solvation data and thus enables validation of simulation data regarding individual Kirkwood-Buff integrals between the molecular species present in the system. Physically the Kirkwood-Buff integrals provide a measure of mutual affinity between the solution components and hence find applications in the studies related to preferential solvations of biomolecules and their effects on the stability on the biomolecules with varying solution components.[8,9]

In this thesis, the Kirkwood-Buff theory and Kirkwood-Buff integrals are applied to modeling and understanding biologically relevant aqueous solutions with the help of molecular dynamics simulations. A new coarse-graining method has been developed using Iterative Boltzmann Inversion method (IBI)[10] combined with Kirkwood-Buff theory which reproduces the molecular solution-structures at pair-level and shows accurate variations of the chemical potentials or activity coefficients of the components of the solutions with varying solution composition. This method, KB-IBI, has been applied to derive single-site coarse-grained models for urea and water, which may be used for the future-studies of urea-denaturations of proteins. Single-site models for small hydrophobic molecule (benzene) in urea-water have also been developed and the salting-in effects of urea on benzene are studied. In addition to developing coarse-grained models for aqueous solutions, Kirkwood-Buff theory has been applied to understand the ion-pairing mechanisms of biologically important anions, such as acetate and dimethyl phosphate, with Hofmeister monovalent cations in water with all-atomistic computer simulations. Explanations for the salting-out effects of the aqueous alkali chlorides on benzene are also reported with the aid of Kirkwood-Buff theory.

This thesis is organized as follows: Chapter 2 serves as a brief introduction to Kirkwood-Buff theory and its relations to thermodynamic quantities relevant to study solvation mechanisms. In Chapter 3, current progresses on the applications of Kirkwood-Buff theory to molecular dynamics simulations are reviewed. The review includes atomistic force-field development using Kirkwood-Buff theory, applications of the theory to understand solution properties and technical aspects regarding the computation of the Kirkwood-Buff integrals. Chapter 4 reports the general procedure to develop coarse-grained models using Kirkwood-Buff theory (KB-IBI) and its application to develop single-site models for urea-water and benzene-water systems. Results show that KB-IBI potentials reproduce the correct variation in the activity coefficients of the solutes and keep molecular pair-structure accurate. KB-IBI data are compared with all-atomistic and experimental data and the limitation in the transferability of the KB-IBI potentials over varying solution concentrations is also discussed. Chapter 5 includes the KB-IBI results for small hydrophobic solute (benzene) in urea-water. The salting-in effects on benzene by different urea concentrations are studied with coarse-grained models where urea-water models are taken from pure binary urea-water mixtures as reported in Chapter 4 and the solute-solute, solute-urea and solute-water potentials are reparametrized using KB-IBI. This report examines a general transferability of the binary urea-water KB-IBI potentials to ternary systems. In addition to that, a cluster analysis of benzene in urea-water with different urea concentrations is performed and reproducibility of the free-energy of cluster formation by KB-IBI potentials with respect to all-atomistic results is tested. Aspects of the representability of the structure-based coarse-grained models at a particular state-point are also discussed for the systems of pure-water, binary mixtures of urea-water and ternary mixtures of benzene in urea-water. The reproducibility of the thermodynamic quantities such as pressure and isothermal compressibility is examined. The convergence of these quantities along with the total pair-potential energy of the system is also studied in detail. In Chapter 6, the ion-pairing mechanisms of monovalent dimethyl phosphate and acetate anions with Hofmeister cations, namely lithium, sodium and potassium, are reported. All-atomistic simulation results are compared with experimental results using Kirkwood-Buff theory. Results show that solvent-shared and solvent-separated ion-pairs dominate over contact ion-pairs and determine the Hofmeister series of the osmotic coefficient for dimethyl phosphate or acetate anions; contradictorily, in alkali chloride solutions contact ion-pairs dominate. The results for acetate and chloride solutions are taken from a previous work

by Hess and van der Vegt.[11] Chapter 7 deals with the salting-out phenomena of benzene with alkali chlorides. All-atom simulation data show that the Hofmeister series for salting-out of benzene is determined by the direct benzene-cation correlations. Indirect ion-ion or ion-water interactions do not play any role to rank the cations in terms of their ability to salt-out benzene. A stronger correlation between lithium and benzene serves as a possible explanation for the anomaly of lithium making it less salting-out agent than sodium and potassium. A geometric packing effect is found to be the cause of benzene-lithium correlations and the partial charges on the benzene molecules do not play any significant role for the benzene-lithium correlations. It is also found out that water-water correlations are not ion-specific. In Chapter 8 a technical account for calculating Kirkwood-Buff integrals is given. Results include the Kirkwood-Buff integrals obtained by integrating the pair correlation functions and by using the method introduced by Schnell and coworkers[12] which calculates the Kirkwood-Buff integrals from the particle number fluctuations in small sub-volumes of the systems. All-atom simulation results for binary mixtures of urea-water and methanol-water show that Schnell's method is more precise than the integration method for smaller systems but both the methods require an equally long sampling time scale for the Kirkwood-Buff integrals to converge. In the last chapter, Chapter 9, a brief summary of the results and possible outlook of this work are included.

Bibliography

[1] Hofmeister, F. *Arch. Exp. Pathol. Pharmakol.* **1888**, *24*, 247.

[2] Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795.

[3] Collins, K. D. *Proc. Natl. Acad. Sci. USA* **1995**, *92* 5553.

[4] Baldwin, R. L.; *Biophys. J.* **1996**, *71*, 2056.

[5] Brini, E.; Algaer, E. A.; Ganguly, P.; Rodríguez-Ropero, F.; Li, C.; van der Vegt, N. F. A. *Softmatter* **2013**, *9*, 2108.

[6] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

[7] Ben-Naim, A. *J. Chem. Phys.* **1977**, *67*, 4884.

[8] Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.

[9] Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Biochem. Biophys.* **2008**, *50*, 1.

[10] Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.

[11] Hess, B.; van der Vegt, N. F. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13296.

[12] Schnell, S. K.; Liu, X.; Simon, J.-M.; Bardow, A.; Bedeaux, D.; Vlugt, T. J. H.; Kjelstrup, S. *J. Phys. Chem. B* **2011**, *115*, 10911.

2 Kirkwood-Buff Theory and Solvation Thermodynamics

The Kirkwood-Buff (KB) theory of solutions is one of the most powerful theories related to solvation thermodynamics and other properties of solutions. The theory introduced by Kirkwood and Buff in 1951 relates thermodynamic properties of solutions to molecular pair-distribution functions of the solutions without making any assumptions regarding the non-covalent interaction operative between the atoms/molecules.[1] This theory is exact, it does not assume any pair-wise additivity of the interaction potentials and it is applicable to any number of molecular species of any type and shape. KB theory uses the integrals of the radial distribution functions (RDFs) between the molecular species present in the solutions and relates them to the thermodynamic properties of the solutions such as compressibility, partial molar volumes or the derivatives of the chemical potentials with changing solution compositions which can also be found from experimental studies. The integrals, named as Kirkwood-Buff integrals (KBIs), can also be calculated alternatively from the cross-fluctuations of the particle-numbers in solutions. Inversion of the Kirkwood-Buff theory, derived by Ben-Naim,[2] relates the thermodynamic quantities of the solutions to the individual KBIs between the molecular species present in the solutions. By this means, experimental thermodynamic results can be analysed on the basis of KBIs and on the other hand KBIs obtained from theoretical or simulations studies can directly be compared with the experimental results. A detailed derivation of the Kirkwood-Buff theory and molecular distribution functions can be found in the literature.[3] In this chapter important thermodynamic relations regarding KB theory and the relations between the KBIs and the thermodynamic quantities of the solutions are presented briefly.

2.1 Kirkwood-Buff Integrals

The Kirkwood-Buff integrals are defined as the integrals of the molecular radial distribution functions over volume in a grand-canonical ensemble (constant chemical potential μ, volume V and temperature T) and given by

$$\mathrm{G}_{ij} = \lim_{R\to\infty} 4\pi \int_0^R \left[\mathrm{g}_{ij}^{\mu VT}(r) - 1\right] r^2 dr, \tag{2.1}$$

where $\mathrm{g}_{ij}^{\mu VT}(r)$ is the RDF and G_{ij} is the KBI between the particle types i and j present in the system. Physically the KBIs reflect the mutual affinities between the particle types where higher positive value of KBI shows higher affinity. For the solutions which are away from their critical points the correlations between the particles are often short-ranged ($R < 1 - 2\,\mathrm{nm}$) and the contributions to the KBIs only come from the fluctuations as determined by the solvation shells. Hence the KBIs can be thought to be a *local* property of a solution and subsequently $\rho_j \mathrm{G}_{ij}$ can be interpreted as the change in the number of particles/molecules of type j in a spherical region of radius R ($R < 1-2\,\mathrm{nm}$) before and after placing a particle/molecule of type i where ρ_j is the number density of particle type j. KBIs are symmetric as $\mathrm{G}_{ij} = \mathrm{G}_{ji}$. Alternatively the KBIs can be calculated from the fluctuations and the cross-fluctuations of the particle numbers in a system using the equation

$$\mathrm{G}_{ij} = V\left[\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_i\rangle}\right] \tag{2.2}$$

where N_i is the number of particles of type i within the system, V is the volume of the system, and δ_{ij} is Kronecker delta. $\langle..\rangle$ denotes the grand-canonical ensemble average.

2.2 KBIs and Thermodynamic Quantities

As mentioned earlier KBIs of a solution can be related to the thermodynamic quantities like compressibility, partial molar volumes, derivative of chemical potentials of the species present in the solution. Here in this section mathematical equations re-

lating the KBIs to these quantities for binary mixtures (solvent(W)+solute(S)) and ternary mixtures (solvent(W)+cosolvent(C)+solute(S)) are presented.

2.2.1 Binary mixtures

For binary mixtures isothermal compressibility (κ_T), partial molar volume of the solute or the solvent ($\bar{V}_S$ or $\bar{V}_W$ respectively), derivative of the chemical potential of the solute with respect to changing solvent concentration and solute concentration (μ_{SW} and μ_{SS} respectively) can be given in terms of the KBIs between solute and solute (G_{SS}), between solute and solvent (G_{SW}) and between solvent and solvent (G_{WW}) as

$$\kappa_T = \frac{\zeta}{k_B T \eta} \ , \tag{2.3}$$

$$\bar{V}_S = \frac{1 + \rho_W \left(\mathrm{G}_{WW} - \mathrm{G}_{SW}\right)}{\eta} \ \text{ and } \ \bar{V}_W = \frac{1 + \rho_S \left(\mathrm{G}_{SS} - \mathrm{G}_{SW}\right)}{\eta} \ , \tag{2.4}$$

$$\mu_{SW} = \left(\frac{\partial \mu_S}{\partial N_W}\right)_{T,p,N_S} = -\frac{k_B T}{V \eta} \ \text{ and } \ \mu_{SS} = \frac{\rho_W k_B T}{\rho_S V \eta} \tag{2.5}$$

where k_B is the Boltzmann constant and η and ζ are given by

$$\eta = \rho_S + \rho_W + \rho_S \rho_W \left(\mathrm{G}_{SS} + \mathrm{G}_{WW} - 2\mathrm{G}_{WS}\right) \tag{2.6}$$

and

$$\zeta = 1 + \rho_S \mathrm{G}_{SS} + \rho_W \mathrm{G}_{WW} + \rho_S \rho_W \left(\mathrm{G}_{SS}\mathrm{G}_{WW} - \mathrm{G}_{SW}^2\right) \ . \tag{2.7}$$

For stable solutions $\eta > 0$ and $\zeta > 0$ as in stable solutions μ_{SS}, μ_{WW}, $\kappa_T > 0$ and $\mu_{SW} < 0$. The chemical potentials of the system can be defined in number density/molar scale (M), molal scale (m) or mole-fraction scale (X) as

$$\begin{aligned} \mu_S &= \mu_S^{0,M} + RT \ln \gamma_S^M M_S \\ \mu_S &= \mu_S^{0,m} + RT \ln \gamma_S^m m_S \\ \mu_S &= \mu_S^{0,X} + RT \ln \gamma_S^X X_S \end{aligned} \tag{2.8}$$

where M_S, m_S and X_S are the concentrations of the solute on the molar, molal and mole-fraction concentration scale; μ^0-s are the chemical potentials of the solute in the limit of zero concentration of solute and the γ_S-s denote the activity coefficients of the solute in different concentration scales with R being the gas constant. Subsequently the derivatives of the chemical potentials with changing solute concentration can also be given in number density/molar scale and in mole-fraction scale which are more relevant in experimental studies.

$$\left(\frac{\partial \mu_S}{\partial \rho_S}\right)_{T,p} = \frac{k_B T}{\rho_S \left(1 + \rho_S \left(\mathrm{G}_{SS} - \mathrm{G}_{SW}\right)\right)} \tag{2.9}$$

$$\left(\frac{\partial \mu_S}{\partial X_S}\right)_{T,p} = k_B T \left(\frac{1}{X_S} - \frac{\rho_W \left(\mathrm{G}_{SS} + \mathrm{G}_{WW} - 2\mathrm{G}_{WS}\right)}{1 + \rho_W X_S \left(\mathrm{G}_{SS} + \mathrm{G}_{WW} - 2\mathrm{G}_{WS}\right)}\right) \tag{2.10}$$

It is interesting to notice that the change in the chemical potential of the solute with changing solute concentration does not depend on the solvent-solvent (often water-water) correlation as G_{WW} does not appear in the Equation 2.9. This is consistent with the known fact the solvent structural changes are enthalpy-entropy compensating in the solvation free-energy. The activity coefficient related to the chemical potential and the free-energy of solvation of the solutes is an important thermodynamic quantity to understand the nonideality in a solution. The derivative of the molar activity coefficient (γ_S^M) with changing solute concentration can also be given by the KBIs as

$$\left(\frac{\partial \ln \gamma_S^M}{\partial \ln \rho_S}\right)_{T,p} = -\frac{\rho_S \left(\mathrm{G}_{SS} - \mathrm{G}_{SW}\right)}{1 + \rho_S \left(\mathrm{G}_{SS} - \mathrm{G}_{SW}\right)} \tag{2.11}$$

where $-k_B T \ln \gamma_S^M$ is the solvation free-energy of the solute.

2.2.2 Ternary mixtures

Similar to the binary mixtures, the thermodynamic quantities like partial molar volume, derivatives of free-energy of solvation of solutes or derivatives of chemical potentials can be expressed in terms of the KBIs in a ternary solution of solute(S), solvent(W) and cosolvent(C). Ternary mixtures are important to study the cosolvent effects on the stability of the solutes. In a biological system conformational changes in biomacromolecules can occur upon changing cosolvent composition or

concentration depending on the relative stability of the solute molecules in different solvent-cosolvent environment. Experimental studies often explain these phenomena in terms of preferential solvation or preferential binding coefficients. In equilibrium dialysis experiments preferential binding coefficients are given by the change in the solvent (water) or cosolvent concentration with changing solute (biomolecule) concentration. The preferential binding coefficient between the solute and the cosolvent can be expressed as [4–8]

$$\Gamma_{SC} = \left(\frac{\partial m_C}{\partial m_S}\right)_{T,\mu_W,\mu_C} \tag{2.12}$$

where m_i-s are the molal concentrations of the species. Physically, preferential binding coefficients signify the relative binding of solute with cosolvent over solvent. If the conformational changes in biomolecules, for example denaturation of protein, are considered as processes with two stages (for proteins, native(N) or denatured(D)), then for the biomolecular processes the change in the equilibrium constant with changing cosolvent concentration at infinite dilution of the solute can be given by

$$\left(\frac{\partial \ln K}{\partial \ln \gamma_C}\right)^0_{T,p,\mu_W} = \Delta B_C - \frac{\rho_C}{\rho_W}\Delta B_W \tag{2.13}$$

where γ_C is the activity coefficient of the cosolvent in any scale and ΔB_C and ΔB_W are the difference in the cosolvent and solvent binding to each state of the process respectively. The infinite dilution condition of the solute is represented by the superscript 0. With the help of the approximation

$$\Gamma_{SC} = \left(\frac{\partial m_C}{\partial m_S}\right)_{T,\mu_W,\mu_C} \approx \left(\frac{\partial m_C}{\partial m_S}\right)_{T,p,\mu_C} \tag{2.14}$$

and using Equation 2.13 it can be shown that

$$\left(\frac{\partial \ln K}{\partial \ln \gamma_C}\right)^0_{T,p,\mu_W} = \Gamma^D_{SC} - \Gamma^N_{SC} = \Delta\Gamma_{SC} = \Delta B_C - \frac{\rho_C}{\rho_W}\Delta B_W \tag{2.15}$$

where D and N are the two stages of the process. It shows that if the cosolvent binds to the solute more in its denatured stage than its native stage then it would

bias the equilibrium towards denatured stage, hence the cosolvent can be termed as denaturant.

Kirkwood-Buff analysis of the ternary solutions at infinite dilution of solute in open systems can relate the preferential binding coefficients to the Kirkwood-Buff integrals between solute and cosolvent and between solute and solvent as

$$\Gamma_{SC} = \rho_C \left(\mathrm{G}_{SC} - \mathrm{G}_{SW}\right) = \Delta N_{SC} - \frac{\rho_C}{\rho_W} \Delta N_{SW} \tag{2.16}$$

where ΔN_{ij} is the excess coordination number of species j around species i relative to the bulk solution. In a closed system it can also be shown that the change in the pseudo chemical potential of the solute,[3] defined as the change in the Gibbs free-energy for placing a particle/molecule at a fixed position in a solution from a fixed position in vacuum, ($\mu_S^* = \mu_S - RT \ln \Lambda_S^3 \rho_S$, Λ is the thermal deBroglie wavelength) with the change in the cosolvent concentration is related to the specific binding coefficient between the solute and the cosolvent through the equation

$$-\frac{1}{k_B T} \left(\frac{\partial \mu_S^*}{\partial \ln \rho_C}\right)_{T,p}^{0} = \frac{\rho_C \left(\mathrm{G}_{SC} - \mathrm{G}_{SW}\right)}{1 + \rho_C \left(\mathrm{G}_{WW} - \mathrm{G}_{CW}\right)} \tag{2.17}$$

and using Equation 2.16 and transforming Equation 2.9 by replacing S with C (from the property of a binary solution of the solvent and the cosolvent, this is valid as the solute is in infinite dilution) we get

$$-\frac{1}{k_B T} \left(\frac{\partial \mu_S^*}{\partial \ln \rho_C}\right)_{T,p}^{0} = \frac{1}{k_B T} \Gamma_{SC} \left(\frac{\partial \mu_C}{\partial \ln \rho_C}\right)_{T,p}. \tag{2.18}$$

For a stable solution $\left(\frac{\partial \mu_C}{\partial \ln \rho_C}\right)_{T,p} > 0$. So the sign of the derivative of the pseudo potential of the solute with increasing cosolvent concentration depends on the sign of the preferential binding coefficient Γ_{SC}. Using Equation 2.16 we find that if $\mathrm{G}_{SC} > \mathrm{G}_{SW}$, *i.e.* if the solute binds more with the cosolvent than it binds to the solvent, then upon increasing the cosolvent concentration the pseudo potential of the solute decreases. This is the condition for "salting-in" of the solute by the cosolvent. The opposite phenomenon where the pseudo chemical potential of the solute increases (the stability of the solute in the solution decreases) with increasing cosol-

vent concentration is called “salting-out”. Similarly the change in the free-energy of solvation of the solute can be given by

$$\left(\frac{\partial \Delta G_S}{\partial X_C}\right)^0_{T,p} = -\frac{RT\left(\rho_W + \rho_C\right)^2}{\eta\prime}\left(\mathrm{G}_{SC} - \mathrm{G}_{SW}\right) \tag{2.19}$$

where $\eta\prime = \rho_W + \rho_C + \rho_W\rho_C\left(G_{WW} + G_{CC} - 2G_{CW}\right) > 0$ for stable solutions. If we define another term $\zeta\prime$, similar to the Equation 2.7 and by replacing S by C as $\zeta\prime = 1 + \rho_S\mathrm{G}_{CC} + \rho_W\mathrm{G}_{WW} + \rho_C\rho_W\left(\mathrm{G}_{CC}\mathrm{G}_{WW} - \mathrm{G}^2_{CW}\right)$ then the isothermal compressibility, partial molar volumes of the solvent and the cosolvent for infinite dilution of the solute take exactly the form of Equation 2.3 and 2.4 as

$$\kappa_T = \frac{\zeta\prime}{k_B T \eta\prime} \ , \tag{2.20}$$

$$\bar{V}_C = \frac{1 + \rho_W\left(\mathrm{G}_{WW} - \mathrm{G}_{CW}\right)}{\eta\prime} \ \text{ and } \ \bar{V}_W = \frac{1 + \rho_C\left(\mathrm{G}_{CC} - \mathrm{G}_{CW}\right)}{\eta\prime} \ , \tag{2.21}$$

and the partial molar volume of the solute can be given by

$$\bar{V}_S = k_B T\kappa_T - \rho_W\bar{V}_W\mathrm{G}_{SW} - \rho_C\bar{V}_C\mathrm{G}_{SC} \, . \tag{2.22}$$

For finite concentration of the solute the derivative of the chemical potential of the solute with respect to the change in the number of cosolvent molecules can be given by

$$\mu_{SC} = \left(\frac{\partial \mu_S}{\partial N_C}\right)_{N_S,N_W,T,p} = \frac{k_B T[1 + \rho_W(\mathrm{G}_{SC} + \mathrm{G}_{WW} - \mathrm{G}_{SW} - \mathrm{G}_{CW})]}{V\eta^{finite}}, \tag{2.23}$$

where η^{finite} is given by

$$\begin{aligned}\eta^{finite} =& \rho_S + \rho_C + \rho_W + \rho_S\rho_C\Delta_{SC} + \rho_C\rho_W\Delta_{CW} + \rho_S\rho_W\Delta_{SW} \\ &- \frac{1}{4}\rho_S\rho_C\rho_W(\Delta^2_{SC} + \Delta^2_{CW} + \Delta^2_{WW} - 2\Delta_{SW}\Delta_{CW} \\ &\qquad - 2\Delta_{SC}\Delta_{SW} - 2\Delta_{SC}\Delta_{CW})\end{aligned} \tag{2.24}$$

with $\Delta_{ij} = \mathrm{G}_{ii} + \mathrm{G}_{jj} - 2\mathrm{G}_{ij}$.

2.2.3 Inversion of the Kirkwood-Buff theory

The Kirkwood-Buff integrals can be used to express different thermodynamic quantities. Inversely, the KBIs can also be expressed in terms of the compressibility, partial molar volume and the derivative of the chemical potential. For a binary mixture of solute and solvent the KBIs between the species can be given as [3]

$$\begin{aligned}
\mathrm{G}_{SS} &= k_B T \kappa_T - \frac{1}{\rho_S} + \frac{\rho_W \bar{V}_W^2 \left(\rho_S + \rho_W\right)}{\rho_S D} \\
\mathrm{G}_{WW} &= k_B T \kappa_T - \frac{1}{\rho_W} + \frac{\rho_S \bar{V}_S^2 \left(\rho_S + \rho_W\right)}{\rho_W D} \\
\mathrm{G}_{SW} &= k_B T \kappa_T - \frac{\bar{V}_S \bar{V}_W \left(\rho_S + \rho_W\right)}{D}
\end{aligned} \tag{2.25}$$

where D is given by

$$D = \frac{X_S}{k_B T} \left(\frac{\partial \mu_S}{\partial X_S} \right)_{T,p} . \tag{2.26}$$

Alternatively the quantity D can also be calculated from the experimental data on the second derivative of the excess Gibbs free-energy of the system with composition or using the data on the partial vapor pressure for ideal gas-mixtures. The inversion theory helps to analyze experimental data on these thermodynamic quantities in terms of the KBIs and the quantities related to the preferential solvation can easily be calculated which find explanations for the stability of solutes in solutions. In this way, simulation and theoretical data can be compared with the experimental data directly. There exists a data bank of the KBIs calculated from the experimental data of variuos binary mixtures. [9] Nevertheless, the inversion to the KB theory finds the individual KBIs between the species, not the individual pair-correlation functions which need to be calculated from simulation or theoretical studies.

Bibliography

[1] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

[2] Ben-Naim, A. *J. Chem. Phys.* **1977**, *67*, 4884.

[3] Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.

[4] Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Biochem. Biophys.* **2008**, *50*, 1.

[5] Scatchard, G. *J. Am. Chem. Soc.* **1946**, *68*, 2315.

[6] Tanford, C. *J. Mol. Biol.* **1969**, *39*, 539.

[7] Wyman, J.; Gill, S. J. *Binding and Linkage*; University Science Books, Mill valley: California, 1990.

[8] Timasheff, S. N. *Biochemistry* **1992**, *31*, 9857.

[9] Wooley, R. J.; O'Connell, J. P. *Fluid Phase Equilib.* **1991**, *66*, 233.

3 Kirkwood-Buff Theory of Solutions and Its Applications to Molecular Simulations of Aqueous Mixtures -a Review

The fluctuation theory of solutions, introduced by Kirkwood and Buff, has been a powerful theoretical tool to study the solvation thermodynamics and is being used in the computer simulations rigorously in the last two decades. Solution properties obtained by computer simulations are directly compared with the experiments by using the Kirkwood-Buff theory and also the solvation mechanisms at the molecular level are provided. In this review we discuss the applications of the Kirkwood-Buff theory to the computer simulations of the aqueous solutions in terms of the force-field development, solution thermodynamics and the cosolvent effects on the solutes using preferential solvations. Also the developments on the technical issues regarding the accurate applications of the Kirkwood-Buff theory to finite-sized systems have been reviewed in details.

3.1 Introduction

The effect of cosolvents on the stability of biomolecules has been a crucially important question since 19th century. Denaturations of proteins by urea or guanidinium chloride[1,2], hofmeister salts[3–5] or the anti-denaturing effects of osmolytes[6] are rigorously being explored, both experimentally and theoretically. In the last century the cosolvent effects have been studied experimentally by means of preferential binding or preferential interactions between the solutes and the solvents and cosolvents.[7–13] Yet, the underlying physical mechanism of the cosolvent effects on biomolecules is not well established. Although, in general, computer simulations of the biomolecular systems can be expected to provide the mechanisms of the conformational changes in biosolutes at the atomic/molecular level, but often simulation studies confront with many challenges such as a) limited access to the time and/or length-scales at which the real processes occur and/or b) validation of the results with experiments. The limitations concerning the time and length-scales can however be approached to overcome by means of increasing computer efficiency or by using more simplified force-field models for the systems such as coarse-grained models.[14,15] But comparing experimental thermodynamic data relevant to solvation mechanisms with the results obtained from simulations is not always straightforward. The fluctuation theory of solutions, which deals with the particle number fluctuations of the solution components and their correlations with the fluctuations of the other components of the solutions, does provide a route to calculate thermodynamic quantities related to solvation from the computer simulations and will be discussed in this review.

The fluctuation theory of solutions, introduced by Kirkwood and Buff in 1951,[16] relates the molecular pair-structures to the quantities important to describe solvation thermodynamics of solutes in stable solutions. This exact theory (KB) does not assume any pair-wise additivity of the interaction potentials of the system and can be applied to any number of solution components of any type and any molecular shape. The KB theory, initially derived for grand canonical systems (μVT), uses the integrals of the pair correlation functions between the solution components, given by[17]

$$\mathrm{G_{ij}} = 4\pi \int_0^\infty \left[\mathrm{g_{ij}^{\mu VT}}(r) - 1\right] r^2 dr, \tag{3.1}$$

where $\mathrm{g}_{\mathrm{ij}}^{\mu \mathrm{VT}}$ is the radial distribution function (RDF) between the solution components i and j and these integrals, termed as Kirkwood-Buff integrals (KBIs) can be related to macroscopic thermodynamic quantities as partial molar volumes, isothermal compressibility and the derivatives of the chemical potentials of the solution components. The physical significance of these KBIs can be viewed as a measure of the mutual affinities between the interacting molecular species in a solution. A higher G_{ij} reflects an over-all stronger chemical attraction between the species i and j (direct or mediated by other components). Alternatively, G_{ij}'s can be expressed in terms of the particle number fluctuations as

$$\mathrm{G}_{\mathrm{ij}} = \mathrm{V}\left[\frac{\langle N_{\mathrm{i}} N_{\mathrm{j}}\rangle - \langle N_{\mathrm{i}}\rangle\langle N_{\mathrm{j}}\rangle}{\langle N_{\mathrm{i}}\rangle\langle N_{\mathrm{j}}\rangle} - \frac{\delta_{\mathrm{ij}}}{\langle N_{\mathrm{i}}\rangle}\right] \tag{3.2}$$

where N_{i} is the number of particles of type i within the system, V is the volume of the system, and δ_{ij} is Kronecker delta. $\langle..\rangle$ denotes the grand-canonical ensemble average. For a ternary system of solute (s), solvent (w) and cosolvent (c) the derivative of the activity coefficient of the solute with the variation in cosolvent concentration (at infinite dilution of the solute) can be given in terms of the KBIs as [17]

$$\lim_{\rho_{\mathrm{s}}\to 0}\left(\frac{\partial \Delta G_{\mathrm{s}}}{\partial x_{\mathrm{c}}}\right)_{p,T} = \frac{RT\left(\rho_{\mathrm{w}}+\rho_{\mathrm{c}}\right)^2}{\eta}\left(\mathrm{G}_{\mathrm{sw}} - \mathrm{G}_{\mathrm{sc}}\right), \tag{3.3}$$

where R is the gas constant, $\eta = \rho_{\mathrm{w}} + \rho_{\mathrm{c}} + \rho_{\mathrm{w}}\rho_{\mathrm{c}}\left(\mathrm{G}_{\mathrm{ww}} + \mathrm{G}_{\mathrm{cc}} - 2\mathrm{G}_{\mathrm{cw}}\right)$, and ρ's are the number densities of the individual components of the solution. A very similar equation can be obtained if one uses the preferential binding coefficients [18] of the solute with solvent and cosolvent. Thus the KBIs manifest a very similar physical sense as the preferential binding coefficients. Theoretical formalisms of KBIs and their relations with thermodynamic quantities, both in open and closed systems, can be found in the works by Ben-Naim and by Smith and co-workers for binary and ternary mixtures [17–20] and also for the solutions with four or more components [21]. From computer simulations, obtaining RDFs is often very straight-forward and by calculating the KBIs from the RDFs one can easily compute the cosolvent effect on the solvation free-energy or other thermodynamic quantities. Of course, the KBIs can also be calculated using equation 3.2; the technical aspects of calculating KBIs are discussed later.

However, the KBIs computed from simulations can be related to thermodynamics and can be compared directly to the experimental data, but as the thermodynamic quantities are given by combinations of different KBIs of a system, one can not be sure about the accuracy of the individual KBIs unless the individual KBIs are compared directly with experiments. The inversion of the Kirkwood-Buff theory, derived by Ben-Naim in 1977,[22] provides the way to calculate the individual KBIs from experimental data. Since then the KB theory has met with an ever-increasing popularity as theoretical analysis of experimental data became possible after the inversion of KB theory. Theoretical works by Ben-Naim[19,23] and later by Pjura and coworkers[24] and Shimizu and coworkers[25,26] used Kirkwood-Buff theory to analyze experimental results in terms of preferential solvations[19], changes in partial molar volumes of the biomolecules[24] or changes in the hydration number around the biomolecules[25]. Other theoretical works to calculate partial molar volumes of the amino acids and polypeptides from Kirkwood-Buff theory were performed by Hirata and coworkers.[27,28]

Apart from the theoretical studies, applications of Kirkwood-Buff theory to computer simulations have become popular since late 1990's because of the robustness of the theory and its ability to calculate thermodynamic data without any approximations and in a computationally cheaper way. Computer simulations together with KB theory provide a molecular picture of the systems which is consistent with experimental results and find applications in force-field development and the studies on preferential solvations for computer-simulated systems. In this review, we discuss the force-fields developed using KB theory till date and applications of the KB theory to biological systems to understand the molecular mechanisms of the systems with the help of computer simulations. Application of the KB theory to computer simulations do come with many technical issues regarding time and length scales of the systems. A technical account for calculating KB integrals, sources of errors while computing KBIs and possible methods for diminishing the errors are also reviewed in this article.

3.2 Kirkwood-Buff Derived Force-fields for Aqueous Solutions

Application of the KB theory to develop thermodynamically consistent atomistic force-fields for simulations of biological systems is introduced by Weerasinghe and Smith[29] by developing an all-atom model for urea for binary mixtures of urea and water where the model is compatible with SPC/E water[30] and the model reproduces experimental KBIs between urea-urea, urea-water and water-water. Since

then several Kikwood-Buff derived force-fields (KBFF) for biologically important molecules are developed by Smith and coworkers (aqueous solutions of acetone,[31] sodium chloride,[32] guanidinium chloride,[33] methanol,[34] amides,[35] salts of polyoxoanions,[36] alkali chlorides;[37] mixtures of thiols, sulphides and disulphides with methanol;[38] mixtures of aromatic amino acids with methanol and water[39]), by van der Vegt and coworkers (aqueous solution of tertiary butanol[40] and alkali chlorides[41]), by Klasczyk and Knecht (alkali chlorides in water[42]) and by Netz and coworkers (aqueous solutions of alkali halides[43] and divalent cations[44]). The main motivation behind KBFF is to reproduce experimental KB integrals which, in many cases, existing other force-fields fail to reproduce and result to a higher self-aggregation of the solute molecules in solutions.[29,41,45,46] In general, to develop KBFF for biomolecules, the bonded and Lennard-Jones parameters are taken from existing force-fields (such as GROMOS[47] or OPLS[48]) and partial charges on the atoms are reparameterized to reproduce experimental KBIs. For more polar atoms Lennard-Jones parameters are also reparameterized and/or scaled with a scaling factor accordingly. Mostly SPC/E[30] water model has been used for KBFF parameterization although SPC[49] or TIP3P[50] water-models provide alternative possibilities to adopt to while parameterizing KBFF. A detailed recipe for KBFF parameterization can be found in the comprehensive reviews by Smith and coworkers.[51,52]

The Kirkwood-Buff derived force-field for urea[29] has been an important contribution in the field of modeling protein denaturations with urea. Many other urea models including OPLS show unphysical urea self-aggregation at higher urea concentrations.[29,53] The KBFF model, which was developed by using the bonded parameters from GROMOS96[47] and tuning the non-bonded Lennard-Jones parameters and the partial Coulomb charges on the atoms, correctly reproduces the derivative of the urea activity coefficient with urea concentration, density and diffusivity of the urea-water binary solutions up to very high urea concentrations ($\approx 8\,m$). The difference in the degrees of urea self-aggregation with OPLS and KBFF model is found to be caused by the difference in the Lennard-Jones parameters whereas the differences in the partial charges do not play any significant role.[54] KBFF urea model in TIP3P water model has successfully been used in folding-unfolding equilibrium studies of protein which provided important insights to the balance of the van der Waals and electrostatic interactions between protein and urea molecules.[55] More recently a single-site coarse-grained force-field for urea-water mixture has been developed by using the structure-based coarse-graining method, Iterative Boltzmann

Inversion (IBI),[56] and the KB theory. The method, named KB-IBI,[57] reproduces the pair-structure of the solution and the variation of the urea activity coefficient with urea concentration correctly with respect to the all-atom simulations. The models which were parameterized for 2 M to 8 M urea concentration also show state-point transferability within fluctuations of 2-3 M urea concentration. Application of these coarse-grained urea-water model to develop coarse-grained models for solutes in urea-water and the state-point representability and transferability are discussed in a later work on KB-IBI.[58]

KBFF all-atomistic ion models for simple electrolytes in water also work better in terms of reproducing experimental preferential solvation coefficients of the salts as the models reproduce experimental KBIs and the derivative of the salt activity coefficients with salt concentrations and the KBFF models do not show unphysical cation-anion aggregations. KBFF models for these electrolytes (halides of alkali and divalent cations) were developed using two different approaches primarily. The first approach[32,37,41] requires tuning of the Lennard-Jones parameters of the ions along with an additional scaling factor for the cation-water van der Waals interactions to balance the too-strong cation-water electrostatic interactions. As the scaling factor used for the cation-water interactions effects the solvation free-energy of the cations, the second approach[43,44] determines the cation-water and anion-water interaction parameters from the single ion-solvation data and then appropriate scaling factors are used for ion-ion van der Waals interactions to reproduce experimental KBIs. Although all of these KBFF ion parameters reproduce the solvation thermodynamics of the binary salt-water mixtures, the applications of these ion models to multicomponent aqueous mixtures are still limited and not well explored. KBFF model for sodium chloride by Weerasinghe and Smith[32] has been used to study the interactions of salts with valine amino acid and the KBFF has been found to be producing very similar results as OPLS ion force-fields in terms of the cation-anion and cation-carboxylate first coordination numbers.[59] The distribution of the ions on protein surfaces for the systems of ribosomal protein, HIV protease and amyloid fibril with aqueous alkali chlorides has been studied[60] using a combination of cation parameters developed by Hess and van der Vegt[41] and chloride parameters developed by Weerasinghe and Smith.[32] Recently salting-out of benzene by aqueous alkali chlorides has been studied using Kirkwood-Buff ion parameters.[61] The study shows that the benzene-ion interactions/correlations depend strongly on the ion parameters and in terms of reproducing experimental benzene salting-out

coefficient the ion-parameters obtained with the second approach of parameterization [43] where the cation-anion interactions are scaled works slightly better than the ion parameters obtained by scaling the water-cation interactions. [37,41]

3.3 Kirkwood-Buff Theory and Solvation Thermodynamics from Computer Simulations

In this section we discuss the application of the KB theory to understand the thermodynamic properties of the aqueous solutions with a focus on the preferential solvation. As mentioned earlier in the Introduction section KB integrals between the solution components bear very similar information as the preferential binding coefficients. For a ternary solution of solute, cosolvent and solvent (water) the change in the solubility of the solute with varying cosolvent concentrations has been a question of immense importance as the conformational changes of macromolecules are very closely related to the change in the solubility of the macromolecules in the solutions. For example, the molecular mechanisms behind the Hofmeister ion-effects [3] on the degrees of the denaturations of the proteins are not yet fully understood. [62] The variation in the solubility of the solutes by cosolvents, referred as salting-in/out effects, can be represented in terms of the variation in the solvation free-energy of the solutes. As KB theory relates the intercomponent total correlations with the derivative of the solvation free-energy of the solutes, the salting-in/out mechanisms of solutes by cosolvents can be studied in terms of the direct or indirect correlations where it involves or does not involve the solute molecules respectively. In Eq. 3.3 the KBIs $\mathrm{G_{sw}}$ and $\mathrm{G_{sc}}$ quantify the direct correlations between solute and water and between solute and cosolvent respectively. Whereas the term η includes the correlations involving solvent and cosolvent only and η is a property of the binary solvent-cosolvent mixtures. Hence η serves as a measure of the indirect correlations that can effect the solvation of the solutes. As the solvation free-energy of solutes in a solvent-cosolvent solution and the term η for binary solvent-cosolvent mixtures can be calculated experimentally, KB theory allows us to verify the simulation results with experiments and predict the correct mechanism by observing the variation in the direct and indirect correlation terms. A molecular dynamics study on the salting-out of benzene by alkali halides in water uses the KB theory and shows the domination of the benzene-cation correlations over the indirect correlations involving the salts and water. [61]

Hydration of small nonpolar solutes and the cosolvent effects on the solutes[63] have been studied with the help of KB theory and Widom particle insertion[64] where aqueous ternary solutions of solutes such as methane or inert gases with cosolvents like urea, guanidinum chloride or chloride salts have been used. It has been shown that the KB theory calculates the change in the solvation free-energy of the solutes more precisely than Widom insertion method and this trend is further observed in later studies also.[40,65] Using KB theory preferential binding of aliphatic hydrocarbons with cosolvents such as urea, dimethyl sulfoxide, sodium chloride, acetone has been studied which also includes the studies on the solvation entropy and enthalpy and the contributions from the solute-solvent interactions to these quantities.[66,67] Effects of salts on the conformational changes of peptides in aqueous solutions have been studied for a system of leucine enkephalin in sodium chloride solution and the preferential ion binding to the peptide has been quantified in terms of the KBIs where direct potential of mean-force approach is not very straight-forward because of the large size of the solute.[68] Along these lines studies on the propane molecules in urea-water,[69] the serum albumin protein and protein-salt interactions,[70] the effect of urea on hen egg white lysozyme,[71] methane in tertiary butanol,[72] alanine peptide in sodium perchlorate solution,[73] Hofmeister ion-effects on amides,[74] the effect of alcohols or urea on small hydrophobes or tetramethyl ammonium ions,[75] tri-glycine in aqueous urea solution[76] have also been reported. Although this review focuses on the computer simulation studies using KB theory but there exist many other theoretical-experimental studies on the solubility and stability of the (macro)molecules in cosolvent mixtures discussed in terms of the KB theory and KB integrals by using the experimental data, which further demonstrates the scope and the importance of the KB theory.[77–82]

For binary solute-solvent mixtures there have been many computer simulation studies on the solute-solute and solute-solvent preferential interactions with the help of KB theory. Apart from the force-field development using KB theory binary aqueous mixtures of urea,[83–85] alcohols,[85–87] sodium perchlorate,[89] sodium sulphate,[90] trifluoroethanol,[91] caffeine,[93] acetamide and N-methylacetamide[92] are also studied with a KB analysis to quantify the solute-solute and solute-solvent preferential interactions. KB theory is used to calculate partial molar volumes of small solutes in binary aqueous mixtures[94,95] and to analyze the effect of pressure on the partial molar volume of protein.[96] In a study of ion-pairing of alkali cations with acetate or chloride anions KB theory is used to relate the intercompo-

nent KBIs to the derivative of the activity coefficient or chemical potential of the ions and the results were compared with the experiment directly.[41] After modifying the interaction parameters suitably and validating the force-fields with experiments the contributions to the KBIs from different ion-pairing mechanisms, such as direct or solvent mediated, have been analyzed. It is found that for acetate solutions ion-pairs mediated by a water molecule dominate over the direct ion-pairs to justify the ion-specific variation in the ion-ion KBIs which determine the ion-specific changes in the activity of the ions, whereas in chloride solutions direct pairing mechanism prevails. This approach of decomposing the KBIs in parts has led to find that water-mediated pairings are more important in phosphate solutions too.[97]

3.4 Computation of Kirkwood-Buff Integrals and Technical Issues

Kirkwood-Buff theory is derived from the particle fluctuations in grand-canonical ensemble. But computer simulations of open systems at constant chemical potential are not straight-forward due to the complications related to the particle insertion.[98] So for the technical convenience most of the simulations are performed in closed canonical isobaric or isochoric ensemble where total numbers of the particles are kept fixed. KBIs ($\mathrm{G_{ij}}$) calculated from a closed system using Eq. 3.1 approach to $-1/\rho_{\mathrm{i}}$ ($\mathrm{i} = \mathrm{j}$) or 0 ($\mathrm{i} \neq \mathrm{j}$) where ρ_{i} is the number density of particle type i. But for the systems with box-sizes much larger than the local correlation lengths between the particles the KBI approaches to a plateau value and usually that value is considered to be the limiting KBI at infinite separation. This approach of calculating the KBIs from Eq. 3.1 for closed systems suffers from many challenges. For closed systems the RDFs do not approach to 1 (typically differ by a term of the order of the inverse of the total particle number[99,100]) and also do not show the correct asymptotic behavior at the tail.[17,101–103] This leads to erroneous plateau values for the KBIs or for the smaller systems does not lead to any plateau value at all. Also the tails of the RDFs suffer from poor convergence due to the poor statistics as it requires longer simulations to move around the masses to longer distances and that gives rise to larger fluctuations in KBIs due to the r^2 term (Eq. 3.1). For aqueous mixtures the convergence issue of the tail of the RDFs is much more complicated for the cosolvents such as urea[53,85,104] or alcohols[85] than for the ions,[41,97] which is due to the proposed microheterogeneity and slow domain-like dynamics of the solutions.[105–108] There have been many attempts to overcome these issues and to calculate more accurate

KBIs from the closed-system simulations and these methods will be discussed in this section.

Calculating the KBIs by taking the average of the volume integrals of the RDFs at a finite range of spatial separations has been the most popular method for computing KBIs from the small systems where the integrals of the RDFs do not approach to any plateau values.[29,31,34,40] The range for the averaging is chosen depending on the system, typically between $\approx 0.9 - 1.0$ to $\approx 1.3 - 1.4$ nm for the aqueous mixtures. But this method does not account for the wrong asymptotic behavior of the tail of the RDF which also has an effect at the shorter distances at ≈ 1.0 nm, which can be observed from the differences in the KBI values obtained from the simulations of smaller and larger systems.[41,85] Also depending on the radial separation one chooses for the averaging of the KBIs, the data may fluctuate significantly.[85] Correction to the tail of the RDF in order to obtain a more pronounced plateau value for the KBIs has been addressed in the literature.[41,102,103,109–113] As mentioned earlier, the asymptote of the tail of the RDF does not converge to 1, rather to $1 - 1/N$ for ideal gas and to $1 - 1/Nk^*$ for pure fluids with presence of interparticle interactions where k^* is the reduced isothermal compressibility given by $k^* = \left(\frac{\partial(\mathrm{P}/K_B\mathrm{T})}{\partial\rho}\right)_{\mathrm{T}}$, k_B being the Boltzmann constant.[103,114,115] For mixtures, the RDFs converge to a similar expression depending on the variation in the density of one species with the change in the chemical potential of the other species at constant temperature.[103] To correct the asymptote of the tail of the RDF, Perera and coworkers have proposed a linear correction to shift the tail of the RDF to 1 at the half of the simulation box length[102] and later the correction is performed by a more complex trigonometric function which does not affect the RDF for the first few peaks in the original function (for aqueous solution this would correspond to first few solvation shells).[103] Van der Vegt and coworkers have corrected the tail of the RDF by considering the correct bulk density of the particles at larger distances[41] and later for all the distances.[85] These two methods, by Perara $et\ al.$ and van der Vegt $et\ al.$ have been compared for urea-water and methanol-water binary mixtures[85] and it has been shown that the later method works better than the previous method to find a well-defined plateau for the KBIs for the smaller simulation boxes, although the accuracy to calculate precise KBIs by these two methods has not been tested. Abildskov and coworkers have proposed another method to correct the integrals of the correlation functions where the integrals are split into three parts, namely direct, indirect and long-range contribution and the indirect correlation part is fitted with an exponential-trigonometric

function where the function approaches 1 faster than r^2 in Eq. 3.1 diverges to infinity.[109] With this method binary mixtures of benzene with methyl acetate and ethanol[109] and systems of normal alkanes[110] have been studied. In a later work with aqueous solutions of alcohols[116] the authors have used the correction method by Verlet[117] and compared it with the methods proposed by Smith and coworkers[29] and by van der Vegt and coworkers.[41,85] It has been found that for the bigger systems the method by van der Vegt $et\ al.$ works better where for the smaller systems Verlet method works equally well or better.

With an alternative approach Vlugt and coworkers have calculated KBIs for small systems embedded in a larger system by using Eq. 3.2 and shown a linear dependency between the KBIs and the inverse of the linear dimension of the small boxes.[118] Earlier using Hill's theory on the statistical mechanics of small systems[119] the authors have deduced a linear relation between the thermodynamic correction factor, as well as molar enthalpy, and the inverse of the linear dimension of small systems in grand-canonical ensemble.[120] The linear relation between the KBIs and the inverse of the linear dimension of the small systems can be extrapolated to the limit of an infinitely large system and the extrapolated values of the KBIs can be considered as the limiting KBIs for the system in which the small systems are embedded, even if the larger system is closed for any particle exchange (canonical ensemble). This method has been applied to calculate KBIs for urea-water and methanol-water mixtures and it has been shown that this method can compute more precise KBIs than the KBIs obtained by the integration of the RDFs and the improvement in the precision of the KBIs becomes more prominent for the smaller system-sizes.[85] Although the time-convergence of the KBIs for aqueous solutions is found to be equally challenging in terms of the simulation length for both the methods. More rigorous theoretical calculations and the relation between the KBIs calculated from these two methods for finite systems can be found in the later works by the authors.[121,122] For aqueous mixture of methanol, direct calculation of the KBIs from the particle number fluctuations of the systems (Eq. 3.2) has also been approached by simulating an all-atomistic system in a larger particle bath with coarse-grained descriptions of the molecules.[87]

Another technical issue concerning the choice of the reference while calculating the KBIs may arise when the system contains larger molecules with higher asphericity. Theoretically KBIs are independent of the choice of the reference center but it has been found that for proteins and peptides convergence of the KBIs improves when the surface of the macromolecule is chosen to be the reference center, more promi-

nently for the smaller system-sizes.[68,123] When the surfaces of the macromolecules are chosen to be the reference, one can only calculate the preferential binding of the macromolecules to the cosolvent over the solvent, $(G_{sw} - G_{sc})$ in Eq. 3.3, but computation of the individual KBIs would not be possible due to the excluded volume contributions to the KBIs. For macromolecules, calculation of the preferential binding coefficients has also been approached by choosing the reference at the center of the nearest residue or group of atoms with respect to the solvent or cosolvent molecules and a non-spherical volume for the normalization of the modified RDFs, namely proximal distributions, has been used.[124] For the aqueous solutions of the salts, calculations of the KBIs encompass a technical issue regarding the electroneutrality condition of the solution. Hence, the cations and the anions cannot be thought as independently fluctuating particles and calculation of the KBIs involving the salt ions is performed by treating the ions indistinguishably.[32,41,97,125] The relation between the KBIs involving salts with indistinguishable ions and the KBIs involving the cations and the anions individually can be found in the literature.[125,126] A method to calculate the individual ion properties from KB theory has been reported where aqueous solution of sodium chloride is simulated.[127]

3.5 Conclusions

Kirkwood-Buff fluctuation theory of the solutions provides a robust theoretical background to study the preferential interactions between the solution components. Computer simulations of aqueous solutions are highly relevant to the biological interests and the KB theory yields a molecular picture of the solvation thermodynamics of solutes in solutions and also offers a simple route to relate the simulation data to the experiments. Several atomistic force-fields have been developed during the last two decades using the KB theory which provide more realistic solution properties in terms of the solvation thermodynamics. KB theory can potentially be applied to study the cosolvent effects in the conformational changes in aqueous solutions,[128] although most of the previous studies have been performed at infinite dilution condition of the solutes as the theoretical treatment for the solutions with finite concentrations of solutes becomes quite rigorous. Technical issues on the convergence of the KBIs have been approached and with the help of the increasing computational power, more precise KBIs are calculated. Fixing the error in the calculation of the KBIs from finite systems with empirical equations has been found to be efficient but the results depend on the functional form one chooses to fix the tail of the RDFs. KB theory

would become more effective and strong in the simulation studies of the solutions if the technical issues are addressed with more accurate theoretical basis.

Bibliography

[1] Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795.

[2] Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5142.

[3] Hofmeister, F. *Arch. Exp. Pathol. Pharmakol.* **1888**, *24*, 247.

[4] von Hippel, P. H.; Schleich, T. *Acc. Chem. Res.* **1969**, *2*, 257.

[5] Collins, K. D.; Washabaugh, M. W. Q. *Quarterly Reviews of Biophysics* **1985**, *18*, 323.

[6] Canchi, D. R.; García, A. E. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273.

[7] Scatchard, G. *J. Am. Chem. Soc.* **1946**, *68*, 2315.

[8] Casassa, E. F.; Eisenberg, H. *Adv. Protein Chem.* **1964**, *19*, 287.

[9] Tanford, C. *J. Mol. Biol.* **1969**, *39*, 539.

[10] Schellman, J. A. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 115.

[11] Wyman, J.; Gill, S. J. *Binding and Linkage*; University Science Books, Mill valley: California, 1990.

[12] Anderson, C. F.; Record, M. T. Jr. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 423.

[13] Timasheff, S. N. *Biochemistry* **1992**, *31*, 9857.

[14] Brini, E.; Algaer, E. A.; Ganguly, P.; Rodríguez-Ropero, F.; Li, C.; van der Vegt, N. F. A. *Softmatter* **2013**, *9*, 2108.

[15] Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750.

[16] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

[17] Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.

[18] Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Biochem. Biophys.* **2008**, *50*, 1.

[19] Ben-Naim, A. *Cell Biophys.* **1988**, *12*, 255.

[20] Smith, P. E. *Biophys. J.* **2006**, *91*, 849.

[21] Kang, M.; Smith, P. E. *J. Chem. Phys.* **2008**, *128*, 244511.

[22] Ben-Naim, A. *J. Chem. Phys.* **1977**, *67*, 4884.

[23] Ben-Naim, A. *J. Chem. Phys.* **1975**, *63*, 2064.

[24] Pjura, P. E.; Paulaitis, M. E.; Lenhoff, A. M. *AIChE J.* **1995**, *41*, 1005.

[25] Shimizu, S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1195.

[26] Shimizu, S.; Boon, C. L. *J. Chem. Phys.* **2004**, *121*, 9147.

[27] Imai, T.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2000**, *112*, 9469.

[28] Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2001**, *114*, 9506.

[29] Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003**, *107*, 3891.

[30] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

[31] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *118*, 10663.

[32] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *119*, 11342.

[33] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2004**, *121*, 2180.

[34] Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005**, *109*, 15080.

[35] Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006**, *27*, 1477.

[36] Zou, J. A Kirkwood-Buff Force Field for polyoxoanions in water *Diss. Kansas State University* 2010.

[37] Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. *J. Chem. Theory Comput.* **2011**, *7*, 1369.

[38] Bentenitis, N.; Cox, N. R.; Smith, P. E. *J. Phys. Chem. B* **2009**, *113*, 12306.

[39] Ploetz, E. A.; Smith, P. E. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18154.

[40] Lee, M. E.; van der Vegt, N. F. A. *J. Chem. Phys.* **2005**, *122*, 114509.

[41] Hess, B.; van der Vegt, N. F. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13296.

[42] Klasczyk, B.; Knecht, V. *J. Chem. Phys* **2010**, *132*, 024109.

[43] Fyta M.; Netz, R. R. *J. Chem. Phys.* **2012**, *136*, 124103.

[44] Mamatkulov, S.; Fyta M.; Netz, R. R. *J. Chem. Phys.* **2013**, *138*, 024505.

[45] Perera, A.; Sokolić, F. *J. Chem. Phys.* **2004**, *121*, 11272.

[46] Auffinger, P.; Cheatham III, T. E.; Vaiana, A. C. *J. Chem. Theory Comput.* **2007**, *3*, 1851.

[47] van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G.; Hochschulverlag AG an der ETH Zürich (1996).

[48] Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

[49] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, 1981; pp 331-342.

[50] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

[51] Weerasinghe, S.; Gee, M. B.; Kang, M.; Bentenitis, N.; Smith, P. E. Developing Force Fields from the Microscopic Structure of Solutions: The Kirkwood-Buff Approach. In *Modeling Solvent Environments*; Feig, M., Ed.; Wiley-VCH: Weinheim, 2010.

[52] Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilib.* **2010**, *290*, 43.

[53] Sokolić, F.; Idrissi, A.; Perera, A. *J. Chem. Phys.* **2002**, *116*, 1636.

[54] Mountain, R. D.; Thirumalai, D. *J. Phys. Chem. B* **2004**, *108*, 6826.

[55] Canchi, D. R.; Paschek, D.; García, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338.

[56] Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.

[57] Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N. F. A. *J. Chem. Theory Comp.* **2012**, *8*, 1802.

[58] Ganguly, P.; van der Vegt, N. F. A. (*submitted*)

[59] Tomé, L. I. N.; Jorge, M.; Gomes, J. R. B.; Coutinho, J. A. P. *J. Phys. Chem. B* **2010**, *114*, 16450.

[60] Friedman, R. *J. Phys. Chem. B* **2011**, *115*, 9213.

[61] Ganguly, P.; Hajari, T.; van der Vegt, N. F. A. (*submitted*)

[62] Lo Nostro, P.; Ninham, B. W. *Chem. Rev.* **2012**, *112*, 2286.

[63] Smith, P. E. *J. Phys. Chem. B* **1999**, *103*, 525.

[64] Widom, B. *J. Chem. Phys.* **1963**, *39*, 2808.

[65] Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2001**, *105*, 11513.

[66] Trzesniak, D.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2004**, *6*, 697.

[67] van der Vegt, N. F. A.; van Gunsteren, W. F. *J. Phys. Chem. B* **2004**, *108*, 1056.

[68] Aburi, M.; Smith, P. E. *J. Phys. Chem. B* **2004**, *108*, 7382.

[69] van der Vegt, N. F. A.; Trzesniak, D.; Kasumaj, B.; van Gunsteren, W. F. *ChemPhysChem* **2004**, *5*, 144.

[70] Shimizu, S.; McLaren, W. M.; Matubayasi, N. *J. Chem. Phys.* **2006**, *124*, 234905.

[71] Kang, M.; Smith, P. E. *Fluid Phase Equilibr.* **2007**, *256*, 14.

[72] Lee, M. E.; van der Vegt, N. F. A. *J. Chem. Theory Comput.* **2007**, *3*, 194.

[73] Asciutto, E. K.; General, I. J.; Xiong, K.; Asher, S. A.; Madura, J. D. *Biophys. J.* **2010**, *98*, 186.

[74] Algaer, E. A.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2011**, *115*, 13781.

[75] Hamsa Priya, M.; Ashbaugh, H. S.; Paulaitis, M. E. *J. Phys. Chem. B* **2011**, *115*, 13633.

[76] Mukherji, D.; van der Vegt, N. F. A.; Kremer, K. *J. Chem. Theory Comput.* **2012**, *8*, 3536.

[77] Matteoli, E.; Lepori, L. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 431.

[78] Zielkiewicz, J. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2925.

[79] García, B.; Aparicio, S.; Alcalde, R.; Leal, J. M. *J. Phys. Chem. B* **2003**, *107*, 13478.

[80] Rösgen, J.; Pettitt, B. M.; Wayne Bolen, D.; *Biophys. J.* **2005**, *89*, 2988.

[81] Shulgin, I. L.; Ruckenstein, E. *Biophys. J.* **2006**, *90*, 704.

[82] Auton, M.; Wayne Bolen, D.; Rösgen, J. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 802.

[83] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *118*, 5901.

[84] Kokubo, H.; Pettitt, B. M. *J. Phys. Chem. B* **2007**, *111*, 5233.

[85] Ganguly, P.; van der Vegt, N. F. A. *J. Chem. Theory Comput.* **2013**, *9*, 1347.

[86] Perera, A.; Sokolić, F.; Almásy, L.; Koga, Y. *J. Chem. Phys.* **2006**, *124*, 124515.

[87] Mukherji, D.; van Vegt N, F. A. der.; Kremer, K.; Delle Site, L. *J. Chem. Theory Comput.* **2012**, *8*, 375.

[88] Gupta, R.; Patey, G. N. *J. Chem. Phys.* **2012**, *137*, 034509.

[89] General, I. J.; Asciutto, E. K.; Madura, J. D. *J. Phys. Chem. B* **2008**, *112*, 15417.

[90] Wernersson, E.; Jungwirth, P. *J. Chem. Theory Comput.* **2010**, *6*, 3233.

[91] Jalili, S.; Akhavan, M. *J. Comput. Chem.* **2010**, *31*, 286.

[92] Lin, B.; Lopes, P. E. M.; Roux, B. Mackerell Jr., A. D. *J. Chem. Phys.* **2013**, *139*, 084509.

[93] Sanjeewa, R.; Weerasinghe, S. *Comp. Theor. Chem.* **2011**, *966*, 140.

[94] Patel, N.; Dubins, D. N.; Pomés, R.; Chalikian, T. V. *J. Phys. Chem. B* **2011**, *115*, 4856.

[95] Patel, N.; Dubins, D. N.; Pomés, R.; Chalikian, T. V. *Biophys. Chem.* **2012**, *161*, 46.

[96] Yu, I.; Tasaki, T.; Nakada, K.; Nagaoka, M. *J. Phys. Chem. B* **2010**, *114*, 12392.

[97] Ganguly, P.; Schravendijk, P.; Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2011**, *115*, 3734.

[98] Beutler, T. C.; van Gunsteren, W. F. *Mol. Simul.* **1994**, *14*, 21.

[99] Hill, T. L. *Statistical Mechanics*; McGraw-Hill: New York, 1956.

[100] Salacuse, J. J.; Denton, A. R.; Egelstaff, P. A.; *Phys. Rev. E* **1996**, *53*, 2382.

[101] Lebowitz, J. L.; Percus, J. K. *Phys. Rev.* **1961**, *124*, 1673.

[102] Perera, A.; Sokolić, F. *J. Chem. Phys.* **2004**, *121*, 11272.

[103] Perera, A.; Zoranić, L.; Sokolić, F.; Mazighi, R. *J. Mol. Liq.* **2011**, *159*, 52.

[104] Stumpe, M. C.; Grubmüller, H. *J. Phys. Chem. B* **2007**, *111*, 6220.

[105] Bowron, D. T.; Finney, J. L.; Soper, A. K. *J. Phys. Chem.* **1998**, *102*, 3551.

[106] Dixit, S.; Crain, J.; Poon, W. C.; Finney, J. L.; Soper, A. K. *Nature* **2002**, *416*, 829.

[107] Allison, S. K.; Fox, J. P.; Hargreaves, R.; Bates, S. P. *Phys. Rev. B* **2005**, *71*, 024201.

[108] Perera, A.; Sokolić, F.; Almasy, L.; Koga, Y. *J. Chem. Phys.* **2006**, *124*, 124515.

[109] Christensen, S.; Peters, G. H.; Hansen, F. Y.; O'Connell, J. P.; Abildskov, J. *Mol. Simul.* **2007**, *33*, 449.

[110] Wedberg, R.; Peters, G. H.; Abildskov, J. *Fluid Phase Equilib.* **2008**, *273*, 1.

[111] Nichols, J. W.; Moore, S. G.; Wheeler, D. R. *Phys. Rev. E* **2009**, *80*, 051203.

[112] Wedberg, R.; O'Connell, J. P.; Peters, G. H.; Abildskov, J. *Mol. Simul.* **2010**, *36*, 1243.

[113] Wedberg, R.; O'Connell, J. P.; Peters, G. H.; Abildskov, J. *J. Chem. Phys.* **2011**, *135*, 084113.

[114] Lebowitz, J. L.; Percus, J. K. *Phys. Rev.* **1961**, *122*, 1675.

[115] Kolafa, J.; Labik, S.; Malijevsky, A. *Mol. Phys.* **2002**, *100*, 2629.

[116] Wedberg, R.; O'Connell, J. P.; Peters, G. H.; Abildskov, J. *Fluid Phase Equilib.* **2011**, *302*, 32.

[117] Verlet, L. *Phys. Rev.* **1968**, *165*, 201.

[118] Schnell, S. K.; Liu, X.; Simon, J.-M.; Bardow, A.; Bedeaux, D.; Vlugt, T. J. H.; Kjelstrup, S. *J. Phys. Chem. B* **2011**, *115*, 10911.

[119] Hill, T. L. *Thermodynamics of Small Systems, Part 1*; Benjamin: New York, 1963.

[120] Schnell, S. K.; Vlugt, T. J. H.; Simon, J.-M.; Bedeaux, D.; Kjelstrup, S. *Chem. Phys. Lett.* **2011**, *504*, 199.

[121] Schnell, S. K.; Vlugt, T. J. H.; Simon, J.-M.; Bedeaux, D.; Kjelstrup, S. *Mol. Phys.* **2012**, *110*, 1069.

[122] Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. *J. Phys. Chem. Lett.* **2013**, *4*, 235.

[123] Baynes, B. M.; Trout, B. L. *J. Phys. Chem. B* **2003**, *107*, 14058.

[124] Paterova, J.; Rembert, K.; Heyda, J.; Kurra, Y.; Okur, H. I.; Liu, W. R.; Hilty, C.; Cremer, P. S.; Jungwirth, P. *J. Phys. Chem. B* **2013**, *117*, 8150.

[125] Friedman, H.; Ramanathan, P. S. *J. Phys. Chem.* **1970**, *74*, 3756.

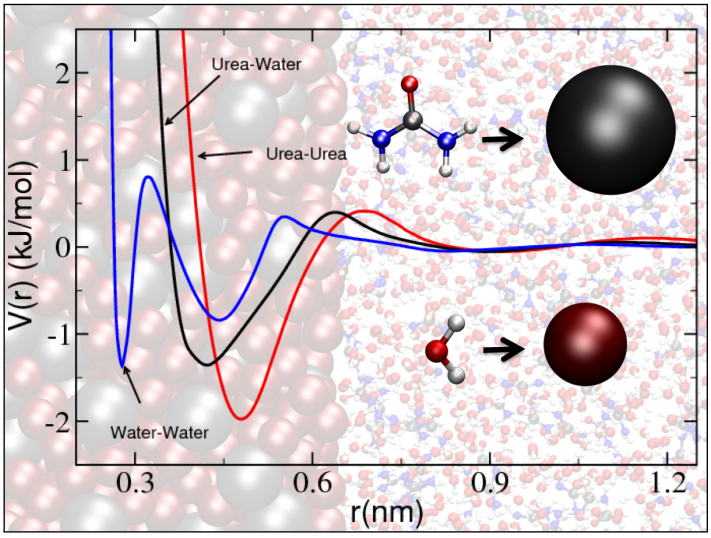
[126] Kusalik, P. G.; Patey, G. N. *J. Chem. Phys.* **1987**, *86*, 5110.

[127] Schnell, S. K.; Englebienne, P.; Simon, J.-M.; Krüger, P.; Balaji, S. P.; Kjelstrup, S.; Bedeaux, D.; Bardow, A.; Vlugt, T. J. H. *Chem. Phys. Lett.* **2013**, *582*, 154.

[128] Smith, P. E.; Matteoli, E.; O'Connell, J. P. *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering, and Biophysics*; CRC Press, 2013.

4 Kirkwood-Buff Coarse-grained Force-fields for Aqueous Solutions

We present an approach to systematically coarse-grain liquid mixtures using the fluctuation solution theory of Kirkwood and Buff in conjunction with the iterative Boltzmann inversion method. The approach preserves, both, the liquid structure at pair level and the dependence of solvation free energies on solvent composition within a unified coarse-graining framework. To test the robustness of our approach, we simulated urea-water and benzene-water systems at different concentrations. For urea-water, three different coarse-grained potentials were developed at different urea concentrations, allowing to simulate urea-water mixtures up to 8 molar urea concentration. In spite of their inherent state point dependence, we find that the single-site models for urea and water are transferable in concentration windows of approximately 2 M. We discuss the development and application of these solvent models in coarse-grained biomolecular simulations.



4.1 Introduction

Biomolecules in water can be salted-in, salted-out or chemically denatured by the presence of cosolvents, such as alcohols, inorganic salts, guanidinium chloride, and urea, to name a few[1,2]. Herein, we are interested in developing systematically coarse-grained (CG) single-site models for water and chemical denaturants such as urea, which may find application in coarse-grained biomolecular simulations[3]. Urea is a well-known salting-in agent: preferential interaction of urea (over water) with nonpolar molecules[4,5] as well as with nonpolar and polar groups on peptides[6,7], including the peptide backbone of proteins[8], favors the solvation of these groups and leads to a decrease of their solvation free energies. A systematic molecular coarse graining method, which provides solvent models that reproduce the solvation free energies while keeping the required structural information, is presently not available and will be proposed in this work.

Effective non-bonded pair potentials for CG models have successfully been developed for polymers[9–12] and nonpolar molecular liquids[13–16] by reversible work techniques in which averages are taken over degrees of freedom no longer represented by the CG model, such as angular orientations. Owing to multi-body correlations, this type of approach will however fail for hydrogen bonded liquids and alternative approaches are needed. Herein, we propose an approach that is based on the thermodynamic theory of Kirkwood and Buff introduced in the early 1950s[17]. Instead of relating the thermodynamic properties to the intermolecular potentials, this theory relates the thermodynamic properties to integrals of radial distribution functions (RDF) over the volume. For solution components i and j, these so-called Kirkwood-Buff integrals (KBI) are defined as[17],

$$\mathrm{G}_{ij} = 4\pi \int_0^\infty [\mathrm{g}_{ij}(r) - 1]\, r^2 dr, \tag{4.1}$$

where $\mathrm{g}_{ij}(r)$ is the RDF and G_{ij} the KBI. Away from the critical point where density fluctuations become long-ranged, contributions to this integral are local and are determined by fluctuations on length scales $R < 1$ nm. Physically, $\rho_j \mathrm{G}_{ij}$ can be interpreted as the change in the number of j molecules in a spherical region of radius R in the solution before and after placing a molecule i at the origin of that region (ρ_j is the number density of component j)[18]. We thus see that G_{ij} is a local quan-

tity which can be used as a measure of the affinity between solution components i and j. In the binary system of cosolvent (c) and water (w) the link to the solvation thermodynamics is given by[18,19],

$$\left(\frac{\partial \ln \gamma_c}{\partial \ln \rho_c}\right)_{p,T} = -\frac{\rho_c \left(\mathrm{G}_{cc} - \mathrm{G}_{cw}\right)}{1 + \rho_c \left(\mathrm{G}_{cc} - \mathrm{G}_{cw}\right)}, \tag{4.2}$$

where $-k_B T \ln \gamma_c$ is the cosolvent solvation free energy (at pressure p, temperature T, and cosolvent number density ρ_c) and γ_c is the cosolvent molar scale activity coefficient. Similar expressions have been derived for systems that have a solute (s) at infinite dilution ($\rho_s \to 0$) in a cosolvent-water solution. In this case, the solvation free energy of the solute (ΔG_s) varies with the solution composition according to[18],

$$\left(\frac{\partial \Delta G_s}{\partial x_c}\right)_{p,T} = \lim_{\rho_s \to 0} \frac{RT \left(\rho_w + \rho_c\right)^2}{\eta} \left(\mathrm{G}_{sw} - \mathrm{G}_{sc}\right), \tag{4.3}$$

where R is the gas constant, $\eta = \rho_w + \rho_c + \rho_w \rho_c \left(G_{ww} + G_{cc} - 2G_{cw}\right)$, and ρ is the number density of individual components of the aqueous solutions. Preferential solvation of the solute by cosolvent molecules ($\mathrm{G}_{sw} - \mathrm{G}_{sc} < 0$) results in a decrease of ΔG_s upon increasing the cosolvent mole fraction x_c (salting-in).

In this paper, we pursue the idea that a CG model provides a good representation of the realistic system if it reproduces the solvent composition dependence of the solvation free energies as expressed by Eq. 4.2 and Eq. 4.3 in a range of nearby concentrations. A conformational transition of a biomolecule driven by changes of the solvation shell composition provides just one illustrative example where this is important. The coarse-grained view provided by Kirkwood-Buff theory tells us that this requirement can be met with models that reproduce the G_{ij}s. Although the RDFs need not necessarily be reproduced to realistically model salting-in and salting-out processes, CG models that represent, both, the RDFs and KBIs of the real system significantly extend the scope and applicability of CG biomolecular simulations.

A CG water-cosolvent model that represents the pairwise liquid structure without sacrificing the required thermodynamic accuracy has previously been reported for benzene in water[20]. The approach reported there however relies on pairwise additivity of hydrophobic interactions between small molecules at low concentration and cannot readily be generalized to hydrophilic compounds. Alternatively, the MARTINI model[21,22] is instead parameterized to reproduce experimental transfer

free energies. Although the MARTINI model is very useful in studies of, amongst others, self-assembly processes, it is not sufficiently accurate to reproduce the dependence of solvation free energies on solvent composition and the corresponding changes in liquid structure. As illustrated by Fig. 4.9 in the Supporting Information, the MARTINI model predicts a typical Lennard-Jones-type fluid structure for an aqueous solution with a polar cosolvent. This structure is however not representative of aqueous systems in which the RDFs show significantly less pronounced long-range oscillations.

4.2 Computational Details

Atomistic simulations were performed with the GROMACS molecular dynamics package.[23] The force field parameters for urea were taken from the Kirkwood-Buff derived force field,[19] for benzene the Gromos 43A1 parameters were used.[24] Water was modeled with the SPC/E potential.[25] The all-atom simulations were performed in the NpT ensemble. The pressure was controlled with a Parrinello-Rahman barostat[26] at 1 atm pressure with a coupling time of 3 ps. The temperature was set to 300 K in all simulations using a Nose-Hoover thermostat[27,28] with relaxation time of 0.5 ps. The integration time step was set to 2 fs and 100 ns trajectories were accumulated. Electrostatic interactions were calculated with the particle mesh Ewald (PME) method.[29] The non-bonded interaction cut-off was chosen as 1 nm. The simulations of aqueous urea were performed with approximately 11000 water molecules, for aqueous benzene the number of water molecules varied between 10000 to 56000. The number of cosolvent molecules were varied according to the concentration. The urea concentrations were taken between 2.6 M and 7.7 M and for benzene between 0.1 M and 0.5 M.

The coarse-grained simulations were performed in a NVT ensemble with the GROMACS simulation package at the average NpT volume of the corresponding atomistic simulation. The equations of motion were integrated using the leap-frog stochastic algorithm. The inverse friction constant was set to 0.2 ps for urea/water and 1.0 ps for the benzene/water systems. The integration time step was set to 4 fs and the cut-off was set to 1.4 nm.

4.3 Results and Discussions

4.3.1 Method and Implementation

CG solvent models that represent the RDFs in principle also represent the thermodynamic solvation properties in Eq. 4.2 and Eq. 4.3. CG methods that optimize effective pair potentials in order to reproduce the RDFs have previously been reported in the literature and include the inverse Monte Carlo [30,31] and iterative Boltzmann inversion (IBI) [32] methods. These methods provide, at least in principle, the required balance of structural and thermodynamic properties, but, as we will show here, need to be further refined, since in practice small variations in the RDFs lead to large variations in the corresponding KBIs owing to the volume integration in Eq. 4.1.

We have simulated all-atom and coarse-grained systems of urea in water and benzene in water. In this work, we use the IBI method implemented in the VOTCA package [33]. The procedure starts from an initial guess for the coarse-grained pair potential, $U_{ij}^{(0)}(r)$, which is obtained from a reference distribution, in this case the RDF, $\mathrm{g}_{ij}^{(ref)}(r)$, between the molecular centers of mass sampled in an all-atom simulation,

$$U_{ij}^{(0)}(r) = -k_{\mathrm{B}}T \ln \mathrm{g}_{ij}^{(ref)}(r). \tag{4.4}$$

The coarse-grained pair potential is iteratively refined until consistency is achieved between the coarse-grained and the reference distributions,

$$U_{ij}^{(n)}(r) = U_{ij}^{(n-1)}(r) + k_{\mathrm{B}}T \ln \left[\frac{\mathrm{g}_{ij}^{(n-1)}(r)}{\mathrm{g}_{ij}^{(ref)}(r)} \right]. \tag{4.5}$$

In every iteration a 10 ns (for benzene) and a 1 ns (for urea) long MD simulation is performed. The final, converged CG potential is then used to run a MD simulation that generates a 25 ns long trajectory. Here, we first start by discussing the results for aqueous urea. Fig. 4.1 shows $\mathrm{g}_{ij}(r)$ and $\mathrm{G}_{ij}(r)$ for urea-urea, urea-water and water-water pairs obtained with the all-atom and IBI coarse-grained models. The KBIs (Eq. 4.1) are obtained from the $\mathrm{G}_{ij}(r)$ functions by taking the limit for large r. Typically, a limiting plateau value is observed in these functions for distances greater than 1 nm, provided that the box dimension is chosen large enough [34]. In this work,

the KBIs are obtained by averaging $\mathrm{G}_{ij}(r)$ in the interval between 1 nm and 1.4 nm. While the RDFs are reproduced within the line thickness, the limiting G_{ij} values are shifted in comparison to the target all-atom G_{ij}s (see black and red curves in Fig. 4.1). In particular in mixtures, any small error in the fitted $\mathrm{g}_{ij}(r)$ at short range can propagate to the tail of $\mathrm{G}_{ij}(r)$, giving rise to the discrepancy observed in Fig. 4.1.

In order to reproduce the exact KBIs, we add a correction term into the coarse-grained potential,

$$\Delta U_{ij}^{(n)}(r) = A(\mathrm{G}_{ij}^{(n)} - \mathrm{G}_{ij}^{(ref)}) \left(1 - \frac{r}{r_{\mathrm{cut}}}\right), \tag{4.6}$$

where $\mathrm{G}_{ij}^{(ref)}$ is the KBI calculated from the reference all-atom simulation and $\mathrm{G}_{ij}^{(n)}$ is the KBI after the n^{th} iteration. The idea behind using the specific ramp can be rationalized as follows: If the KBIs of the CG model are larger than the all-atom KBIs (as in the urea-urea and water-water KBIs in Fig. 4.1), this infers an unphysical local excess coordination of molecules. Therefore, some repulsion needs to be added to the potential in order to weaken the aggregation. Similarly, if the local aggregation is underestimated (as in urea-water KBI in Fig. 4.1), then some attraction is needed in the potential. In principle, we can choose any appropriate functional form. However, we chose to use the simplest function that has been shown to work perfectly well in the case of linear pressure correction.[32] The pre-factor A is system specific and can be tuned based on convergence. With our system and specific simulation protocol, a good estimate of A is in the range between 0.01 and 0.10 kJ nm^{-3} mol^{-1}. The prefactors are summarized in the Supporting Information. The KBIs are significantly improved by using the correction in Eq. 4.6 as a ramp (see the blue curve in the Fig. 4.1). The models obtained in this way are referred to as KB-IBI. The modifications to the VOTCA package[33] used for the KB-IBI method will be a part of the VOTCA release 1.3 and we will also include an example from this work. In the Fig. 4.2, we show a comparative plot of the pairwise coarse-grained potentials obtained from the two separate approaches.

4.3.2 Applications

Having validated the approach for one concentration of urea, we now want to test if the same approach can also be used for a wider range of cosolvent concentrations. Fig. 4.3 shows G_{ij} as a function of urea molar concentration c_u. The unmodified IBI

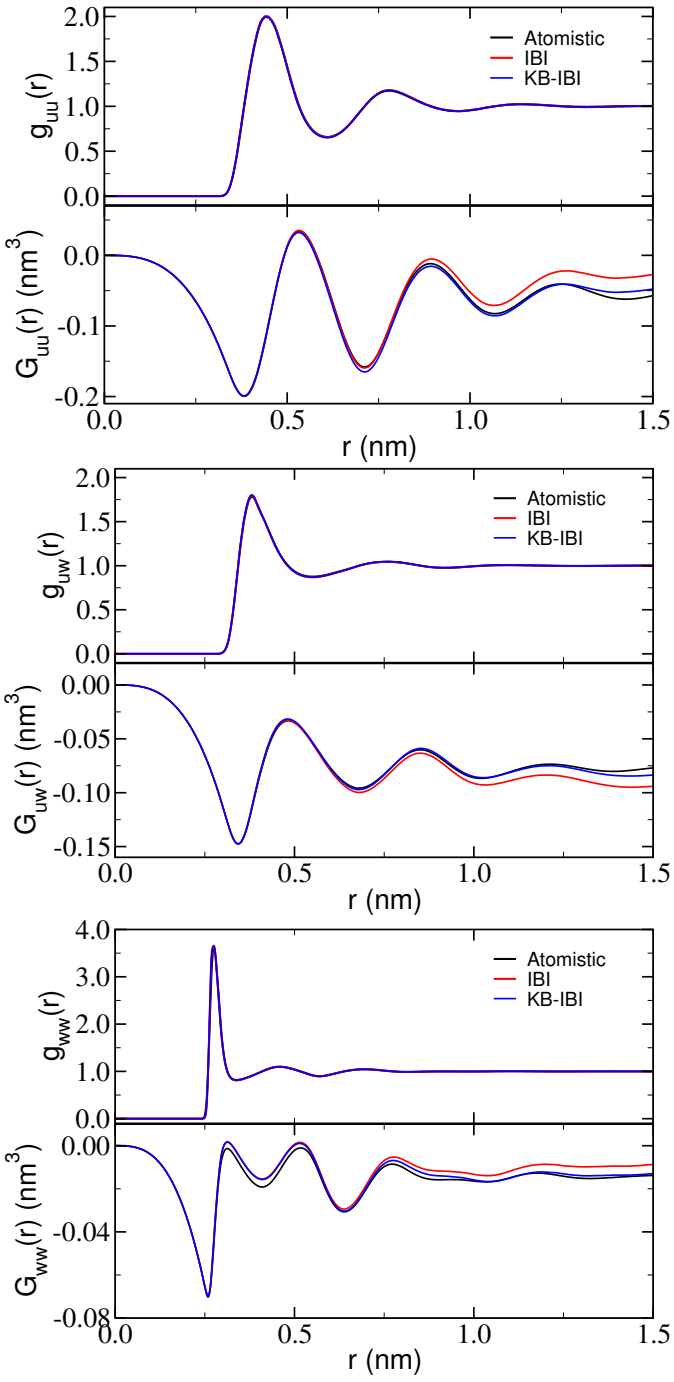


Figure 4.1: Radial distribution functions $g_{ij}(r)$ and running integrals $\mathrm{G}_{ij}(r) = 4\pi \int_0^r [g_{ij}(s) - 1]\, s^2 ds$ for aqueous urea mixture at 4.7 M urea. Comparative data are shown for all-atom and for the IBI and Kirkwood-Buff IBI (KB-IBI) coarse-graining methods. We present data for all three pairs, urea-urea (uu), urea-water (uw) and water-water (ww).

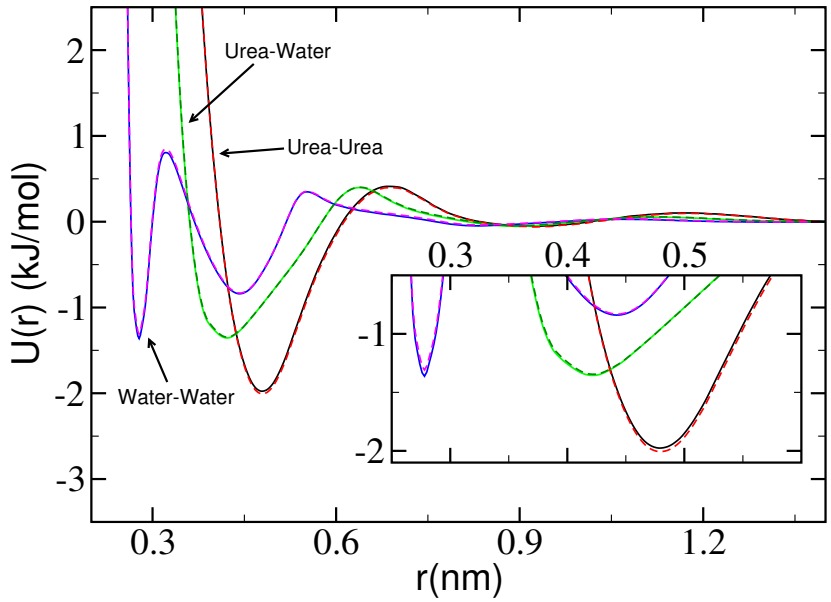


Figure 4.2: Coarse-grained potentials for urea-water mixture obtained at 4.7 M urea solution. Solid lines denote IBI potentials; the dashed lines denote KB-IBI potentials. The inset shows an enlarged view of the potential minima.

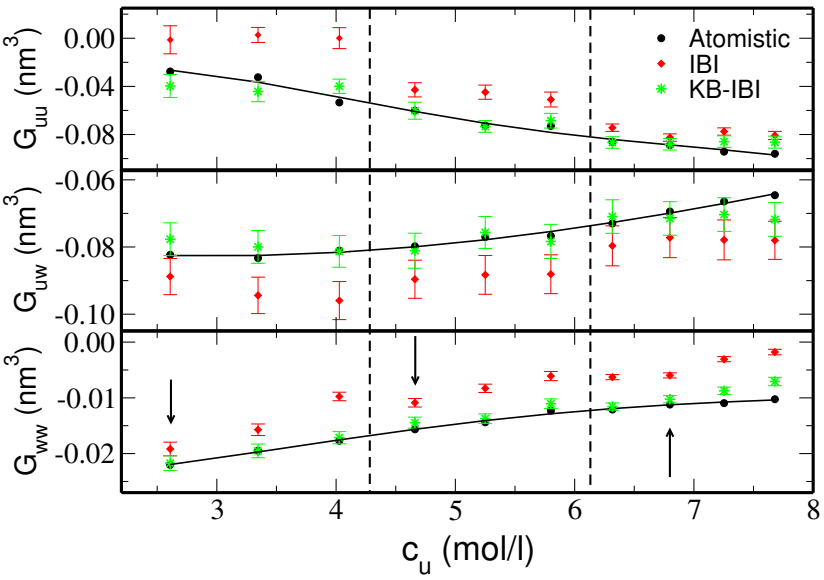


Figure 4.3: Kirkwood-Buff integrals G_{ij} for aqueous urea solutions as a function of the molar urea concentration c_u. Results are shown for all three pairs; urea-urea (uu), urea-water (uw) and water-water (ww). The CG pair potentials (IBI and KB-IBI) were developed for the solution systems (at urea concentration c_u) indicated by the arrows and were subsequently used in the concentration windows bounded by the vertical dashed lines. Solid lines are fits to the all-atom data.

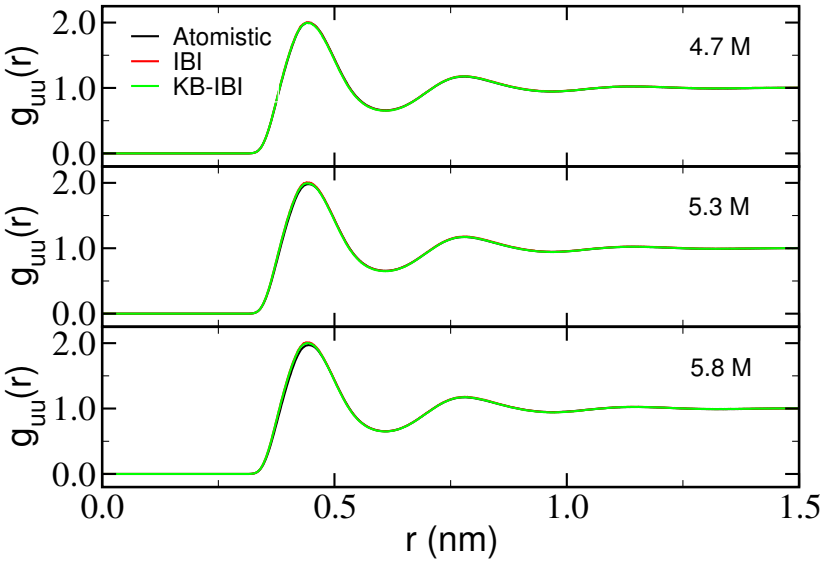


Figure 4.4: Radial distribution function between urea molecules for three different urea molar concentrations. The coarse-grained pair potential, developed for the 4.7 M urea solution, was used for all three concentrations.

model shows significant deviations from the all-atom data. The KB-IBI model does a much better job and reproduces the KBIs and, therefore, the solution thermodynamic properties of the parent atomistic model. We point out that procedures which *identically* match the liquid structure of the CG and all-atom models necessarily provide identical KBIs, but not vice versa. Therefore, it is imperative to investigate to what extent the KB-IBI procedure preserves the liquid structure. In Fig. 4.4, we show the urea-urea RDF for three concentrations corresponding to the middle panel of Fig. 4.3. These results confirm that the RDF is reproduced very well. The urea-urea RDF converges slowest owing to the smaller number of urea molecules compared to water molecules present in the system. The urea-water and water-water RDFs (not shown) show equally good agreement.

The urea and water CG potentials are state point dependent. Their transferability to systems with varying urea concentration is therefore not guaranteed. The CG urea-water systems were however simulated with the same CG potential in a finite window of urea concentrations delineated by the vertical dashed lines in Fig. 4.3. We thus developed the CG potential at only three urea concentrations, namely 2.6 M, 4.7 M and 6.8 M, indicated by the arrows in the Fig. 4.3. The potentials are shown in the Supporting Information. The data in Fig. 4.3 show that the KB-IBI models are transferable in concentration ranges of approximately 2 M. This is partially due to the invariance of the pair structure over such a small concentration range.

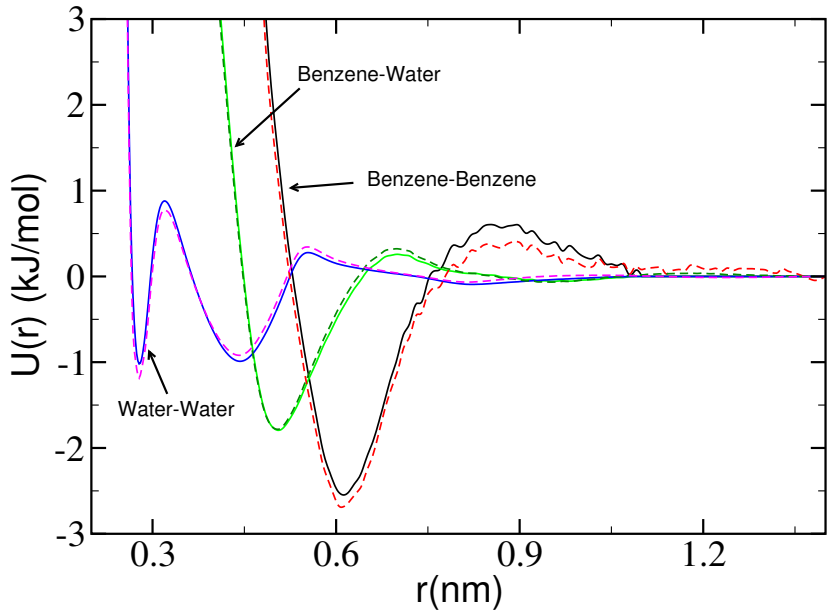


Figure 4.5: Coarse-grained potentials for benzene-water mixture obtained at 0.2 M benzene solution. Solid lines denote IBI potentials; the dashed lines denote the KB-IBI potentials.

Next, we show the results for aqueous benzene solutions. The pairwise CG potential for the benzene-water solution is shown in the Fig. 4.5. Fig. 4.6 presents the G_{ij}s as a function of benzene molar concentration. In this case, the results are significantly better for the KB-IBI model compared to the normal IBI model, which shows too strong benzene-benzene aggregation at all concentrations. The CG potential, developed at $c_b = 0.2$ M, shows good transferability up to a concentration slightly above 0.5 M where the system becomes unstable (phase separates) in the all-atom simulation. In CG simulations with the IBI, KB-IBI, and earlier developed models[20], the benzene-water system however remains stable above 0.5 M. This observation indicates that KB-IBI CG models indeed realistically describe thermodynamically stable solutions, but fail to describe systems outside equilibrium and processes including lipid self-assembly for which the MARTINI model[21] provides a better choice. Fig. 4.7 presents benzene-benzene RDFs. The data clearly supports that the structure is reasonably well reproduced. A closer inspection of the plot reveals that the first peak at 0.6 nm is better reproduced for the lower molar concentrations. However, at larger concentrations, the agreement is relatively poor. Furthermore, longer ranged correlations (beyond 0.75 nm) are always better obtained by the KB-IBI model. We show only the benzene-benzene RDF, which is mostly affected because of poor statistics. The benzene-water and water-water RDFs

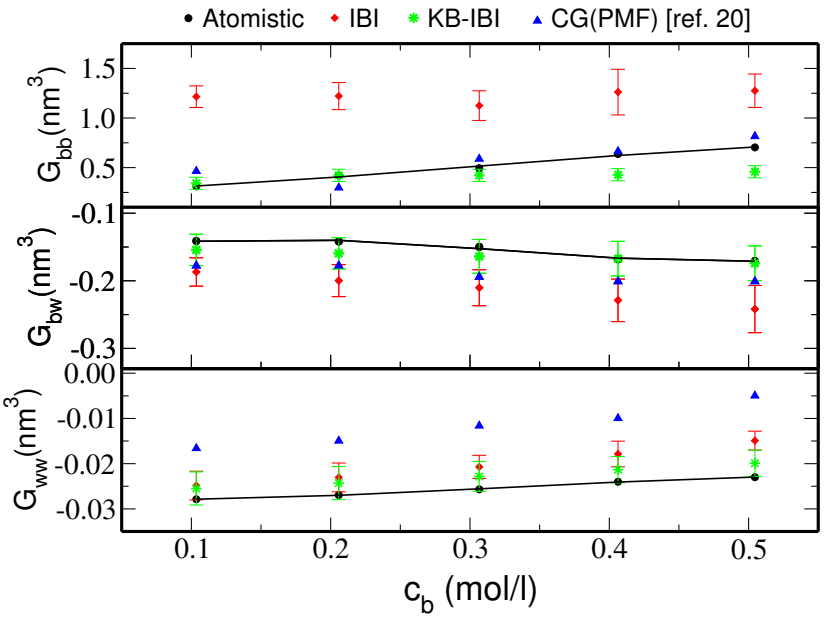


Figure 4.6: Kirkwood-Buff integrals G_{ij} for aqueous benzene solutions as a function of molar benzene concentration c_b. Results are shown for all three pairs; benzene-benzene (bb), benzene-water (bw) and water-water (ww). The CG pair potentials (IBI and KB-IBI) were optimized for the system with $c_b = 0.2$ M and have been used in the CG MD simulations at all concentrations. The CG(PMF) data obtained by Villa et al.[20] are included for comparison. Solid lines are fits to the all-atom data.

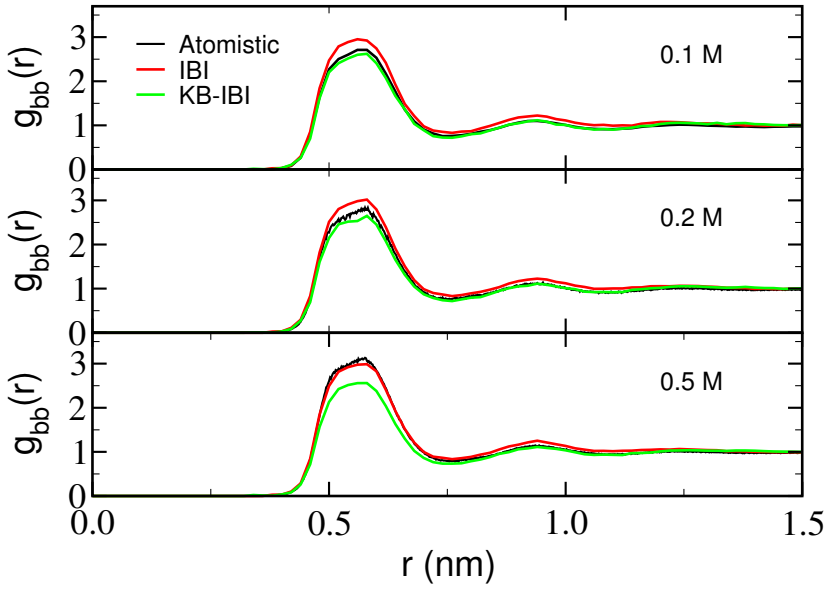


Figure 4.7: Radial distribution function for three different benzene molar concentrations. The coarse-grained potential was optimized for the 0.2 M benzene solution and was used at all three concentrations.

are in closer agreement with the reference atomistic model (not shown). Interestingly, KBI-IBI provides a better converged benzene-benzene RDF with significantly fewer iterations (in total 50 iterations were performed) compared to the standard IBI method (100 iterations).

Finally, we consider the solvation free energies of the aqueous mixture components. The dependence of the urea and benzene solvation free energies on their molar concentration in solution is described by the quantity $f_{cc} \equiv \left(\frac{\partial \ln \gamma_c}{\partial \ln \rho_c}\right)_{p,T}$ and is presented in Fig. 4.8 as a function of the cosolvent (urea or benzene) concentration c_c. The KB-IBI model shows significantly better agreement with the all-atom data and with experiments in comparison to the IBI model.

Practical application of the models discussed in this paper requires further development of pair potentials that describe urea and water interactions with chemical groups of dissolved solutes (peptides, proteins, *etc.*). The KB-IBI method can readily be extended to systems with additional solutes and therefore may provide a useful route to construct non-bonded potentials that can be used in CG simulations of salting-in and salting-out phenomena, protein denaturation or stabilization by chemical denaturants and/or osmolytes. In these applications, the phenomena of interest are driven by fluctuations in solvent composition and CG models must therefore reproduce the changes in solvation free energies with quantitative accuracy, *i.e.;* the CG models must reproduce the G_{ij}s in a suitable range of concentrations.

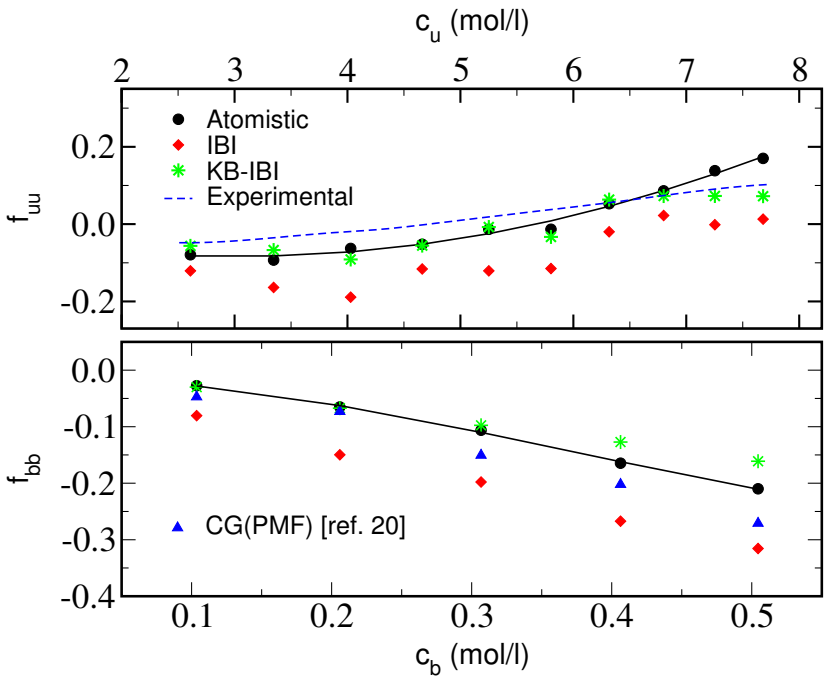


Figure 4.8: Derivative of molar activity coefficient f_{cc} as a function of molar concentration c_c of cosolvent (suffix u for urea and b for benzene). In the case of aqueous urea mixtures the experimental data is taken from Ref.[19]. Black solid lines are fits to the all-atom data.

4.4 Conclusions

We have proposed a systematic molecular coarse-graining approach, which, inspired by Kirkwood-Buff (KB) solution theory, provides a new route to developing CG models that reproduce the liquid structure at pair level and the solvation free energies of the mixture components. The approach is based on the iterative Boltzmann inversion (IBI) method and is denoted KB-IBI. We developed three different coarse-grained KB-IBI potentials for urea/water systems at different urea concentrations, allowing to simulate urea-water mixtures up to 8 molar urea concentration. In spite of their inherent state point dependence, we find that the single-site models for urea and water are transferable in concentration windows of approximately 2 M. We furthermore find that the KB-IBI method provides converged potentials for the liquid mixtures studied here with significantly fewer iterations compared to standard IBI. The KB-IBI method can be easily generalized to multicomponent systems, offering opportunities to parameterize CG non-bonded potentials for interactions between solvent components and chemical groups on biomolecules. A next step in this direction would be to consider single solutes in urea-water mixtures and opti-

mize only the KB-IBI solute-solvent potentials without further optimizing the KB-IBI solvent-solvent potentials, which are then taken from the binary urea-water mixtures studied in this work. Since the salting-in (or salting-out) behavior in dilute solute/urea/water systems is determined by the solute-solvent Kirkwood-Buff integrals only (see Eq. 4.3), while the solvent-solvent KBIs can be assumed to remain unaffected in the limit of very low solute concentration, KB-IBI solute-solvent potentials can be developed for a variety of solutes in combination with a fixed set of potentials for the solvent-solvent interactions.

4.5 Supporting Information

4.5.1 Pre-factors "A" used in the ramp potential

Before applying the Kirkwood-Buff ramp potential in the iterations, a number of standard IBI iterations have been performed. In this way, the KB-IBI iterations start with already a good estimate of the RDFs. The KB-IBI procedure was preceded by 15 standard IBI iterations for the benzene-water systems. The RDFs obtained from the standard IBI potentials do not significantly improve with further IBI iterations but can be improved significantly with further KB-IBI iterations (Fig 4.7, middle panel). The KB-IBI procedure was preceded by 60 standard IBI iterations for the urea-water systems. The A-values that were used as the pre-factors in the ramp potentials (Eq. 4.6) are summarized in Tables 4.1, 4.2, and 4.3.

4.5.1.1 Urea-Water (parameterization at 4.7 M)

Iteration number	**A** $(\mathbf{kJ\ nm^{-3}\ mol^{-1}})$
$1-10$	0.1247
$11-20$	0.0312
$21-50$	0.0062

Table 4.1: A-values used for urea-urea, urea-water and water-water interactions

4.5.1.2 Benzene-Water (parameterization at 0.2 M)

Iteration number	**A** $(\mathbf{kJ\ nm^{-3}\ mol^{-1}})$
$1-10$	0.01247
$11-20$	0.00312
$21-40$	0.00062

Table 4.2: A-values used for benzene-benzene interactions

Iteration number	**A** $(\mathbf{kJ\ nm^{-3}\ mol^{-1}})$
$1-10$	0.01247
$11-20$	0.00312
$21-40$	0.03118

Table 4.3: A-values used for benzene-water and water-water interactions

4.5.2 Urea-Water simulated with MARTINI model

The MARTINI force-fields [21,22,35] do not contain urea parameters. The polar MARTINI $\mathrm{P_5}$ bead most closely mimics urea and has been used here to simulate a 6 molal urea-water mixture. To describe the water beads, the non-polarizable [22] and polarizable [35] MARTINI water models were used. 10 ns long trajectories were obtained at constant temperature of 300 K and constant pressure of 1 bar. The radial distribution functions (RDFs) obtained between the P5 beads and the corresponding Kirkwood-Buff integrals are compared with atomistic urea-urea RDF and corresponding KB integrals in Fig. 4.9.

The MARTINI models show an overemphasized oscillating fluid structure (top panel), which is a characteristic for Lennard-Jones fluids, but it is quite unphysical for water-cosolvent mixtures. The oscillating liquid structure results in running integrals G(r) (lower panel) with long-range oscillations and therefore ill-defined KBIs. A rough estimate of the KBIs, obtained by taking the average over few oscillations of G(r), is a factor 2 or 3 lower than the atomistic urea-urea KBI.

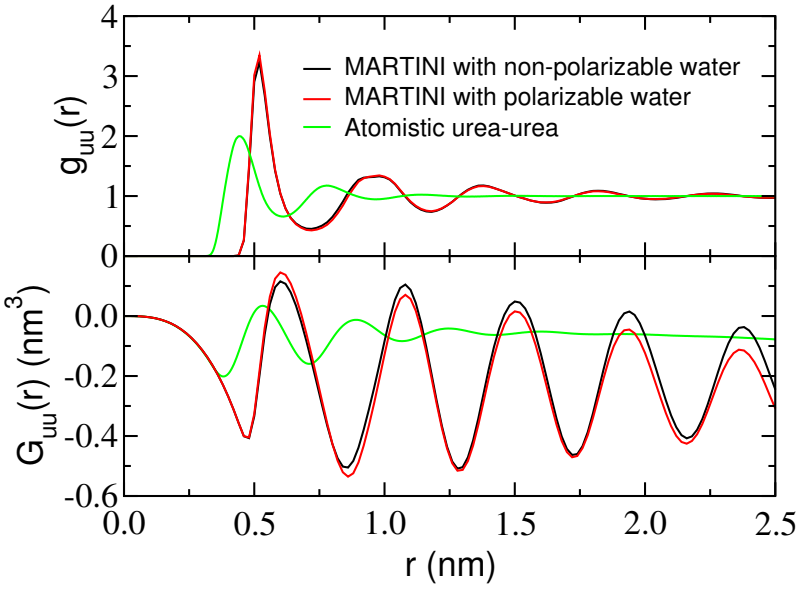


Figure 4.9: Radial distribution functions g_{uu} between P_5 MARTINI beads (upper panel) and corresponding running integrals G_{uu} for a 6 molal cosolvent-water system simulated using two different MARTINI water models [22,35] (black and red lines). For comparison the corresponding urea-urea radial distribution function and running integrals are also shown (green lines).

4.5.3 KB-IBI potentials for urea-water at different urea concentrations

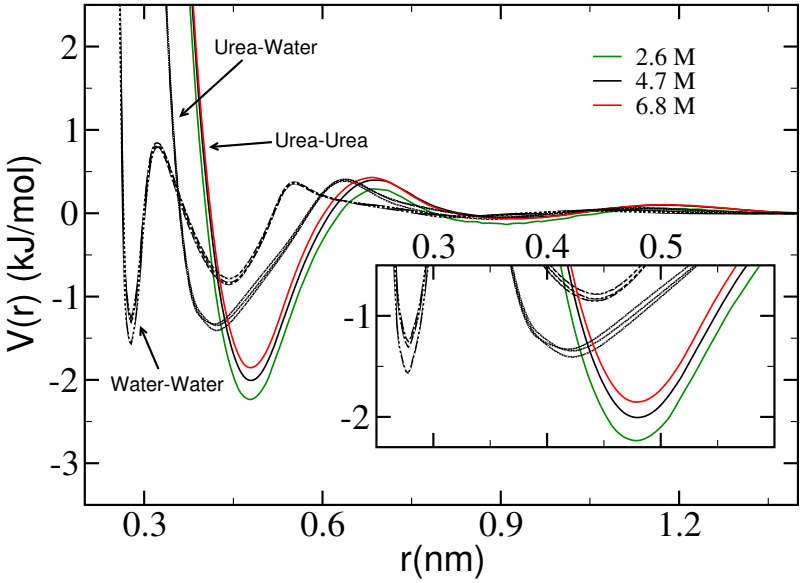


Figure 4.10: Urea-urea, urea-water and water-water KBI-IBI potentials for urea-water mixtures parameterized at three different urea concentrations. Inset shows an enlarged view of the potential minima.

4.6 Acknowledgements

The authors are grateful to Christine Peter for stimulating discussions. We thank Kurt Kremer and Emiliano Brini for useful comments on the manuscript.

Bibliography

[1] Record M. T.; Zhang W. T.; Anderson C. F. *Adv. Protein. Chem.* **1998**, *51*, 281.

[2] Zhang, Y. J.; Cremer, P. S. *Ann. Rev. Phys. Chem.* **2010**, *61*, 63.

[3] Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems* (CRC Press, Boca Raton, 2009).

[4] Trzesniak, D.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2004**, *6*, 697.

[5] Lee, M. E.; van der Vegt, N. F. A. *J. Am. Chem. Soc.* **2006**, *128*, 4948.

[6] Stumpe, M. C.; Grubmüller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126.

[7] Horinek, D.; Netz, R. R. *J. Phys. Chem. A* **2011**, *115*, 6125.

[8] Canchi, D. R.; Paschek, D.; García, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338.

[9] Fukunaga, H.; Takimoto, J.; Doi, M. *J. Chem. Phys.* **2002**, *116*, 8183.

[10] Fritz, D.; Harmandaris, V. A.; Kremer, K.; van der Vegt, N. F. A. *Macromolecules* **2009**, *42*, 7579.

[11] Delle Site, L.; Holm, C.; van der Vegt, N. F. A. *Top. Curr. Chem.* **2012**, *307*, 251.

[12] Li, C.; Shen, J.W.; Peter, C.; van der Vegt, N. F. A. *Macromolecules* **2012**, *45*, 2551.

[13] McCoy, J. D.; Curro, J. G. *Macromolecules* **1998**, *31*, 9362.

[14] Zacharopoulos, N.; Vergadou, N.; Theodorou, D. N. *J. Chem. Phys.* **2005**, *122*, 244111.

[15] Mognetti, B. M.; Virnau, P.; Yelash, L.; Binder, K.; Müller, M.; MacDowell, L. G. *J. Chem. Phys.* **2009**, *130*, 044101.

[16] Brini, E.; Marcon, V.; van der Vegt, N. F. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10468.

[17] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

[18] Ben-Naim, A. *Molecular Theory of Solutions* Oxford University Press: New York, 2006.

[19] Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003**, *107*, 3891.

[20] Villa, A.; Peter, C.; van der Vegt, N. F. A. *J. Chem. Theory Comp.* **2010**, *6*, 2434.

[21] Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750.

[22] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812.

[23] Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, *7*, 306.

[24] van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Hochschulverlag AG an der ETH Zürich (1996).

[25] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

[26] Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.

[27] Nose, S. *Mol. Phys.* **1984**, *52*, 255.

[28] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

[29] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

[30] Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730.

[31] Soper, A. K. *Chem. Phys.* **1996**, *202*, 295.

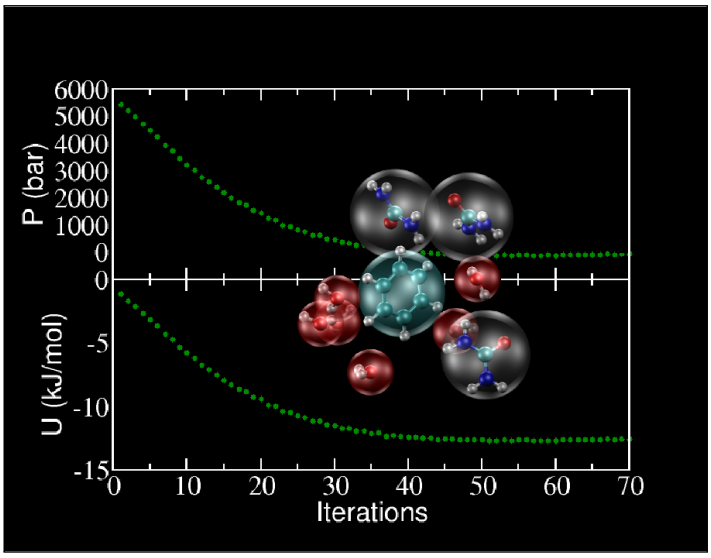
[32] Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.

[33] Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theo. Comput.* **2009**, *5*, 3211.

[34] Schnell, S. K.; Liu, X.; Simon, J. M.; Bardow, A.; Bedeaux, D.; Vlugt, T. H. J.; Kjelstrup, S. *J. Phys. Chem. B* **2011**, *115*, 10911.

[35] Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. *PLoS Comput. Biol.* **2010**, *6*, e1000810.

5 Representability and Transferability of Kirkwood-Buff Iterative Boltzmann Inversion Models for Multicomponent Aqueous Systems



We discuss the application of the Kirkwood-Buff Iterative Boltzmann Inversion (KB-IBI) method for molecular coarse-graining (Ganguly *et al. J. Chem. Theory Comput.* **2012**, *8*, 1802) to multicomponent aqueous mixtures. Using a fixed set of effective single-site solvent-solvent potentials previously derived for binary urea-water systems, solute-solvent and solute-solute KB-IBI coarse-grained potentials have been derived for benzene in urea-water mixtures. Preferential solvation and salting-in coefficients of benzene are reproduced in quantitative agreement with the atomistic force field model. The transferability of the coarse-grained models is discussed and it is shown that free energies of formation of hydrophobic benzene clusters obtained from simulations with the coarse-grained model are in good agreement with results obtained from all-atom simulations. The state-point representability of the coarse-grained models is discussed with respect to reproducing thermodynamic quantities such as pressure, isothermal compressibility and preferential solvation. Combined use of KB-IBI and pressure corrections in deriving single-site coarse-grained models for pure-water, binary mixtures of urea and water and ternary mixtures of benzene in

urea-water at infinite benzene dilution provides an improved scheme to representing the atomistic pressure and the preferential solvation between the solution components. It is also found that the application of KB-IBI leads to a faster and improved convergence of the pressure and potential energy compared to the Iterative Boltzmann Inversion method.

5.1 Introduction

Molecular simulations with detailed atomistic models frequently face limitations in sampling condensed phase systems on sufficiently long time and length scales. While on the one hand chemical specificity is often required, computational efficiency and speed are required on the other hand. Amongst many other examples, a process such as a conformational transition of a solvated macromolecule or protein driven by variations in the thermodynamic activity of (co)solvent components clearly calls for models that are simple yet specific. While coarse-grained (CG) models may potentially bridge several orders in time and length scales, it remains a significant challenge to develop representative CG models for complex molecular systems that are sufficiently transferable such that they can be used in multiscale simulations of soft matter systems under equilibrium and non-equilibrium conditions. Clearly, these types of questions require CG models that reproduce some of the structural and thermodynamic properties of the atomistic system, a requirement which is not straightforwardly met in general.[1–4] While systematic (bottom up) coarse-graining approaches to develop structurally consistent and transferable CG models for molecular liquids and macromolecules exist in the literature,[5–13] significantly less attention has been devoted to deriving CG potentials for molecular mixtures, in particular for aqueous solution systems that contain cosolvent components. Systematic coarse-graining approaches that include preferential solvation in the parameterization procedure, in addition to other structural and thermodynamic properties, may provide new routes to modeling these complex systems at a CG level.

In recent years, the Iterative Boltzmann Inversion (IBI) method[14,15] has gained significant popularity as it provides a robust tool to derive structure-based CG models for condensed phase systems. The method iteratively improves an effective pair potential until agreement is found between the pair correlation functions at the atomistic and CG level of modeling. The derived potential is in principle unique, as stated by the Henderson theorem,[16] which shows that two different pair potentials that generate the same radial distribution function (RDF) can differ by a constant only. In practical applications of IBI it has however been shown that two very different pair potential functions can produce two very similar RDFs which are indistinguishable within very small error margins.[17] Although this suggests that the IBI method is an ill-posed inverse mathematical problem, it provides some additional flexibility to optimize the effective pair potentials in such a way that not only target RDFs are

reproduced but also other thermodynamic “target quantities" such as the energy or pressure, which are both sensitive to variations in the tails of the effective pair potentials.[15] It should be noted though that different thermodynamic targets cannot always be satisfied independently thus requiring a “best compromise" to be found. For example, it is difficult (if not impossible) to match, both, the pressure and the isothermal compressibility of liquid water with single-site IBI models.[18] For more simple systems with Lennard-Jones interactions only, it has been found that by using a minimization procedure that matches the RDF and the pressure an improvement is achieved in reproducing the isothermal compressibility.[19] For solution mixtures, these aspects have so far not been addressed in the literature.

In this paper, we discuss some aspects of the representability and transferability of IBI-based CG models for molecular mixtures with emphasis on aspects of preferential solvation and corresponding thermodynamic changes that lead to salting-in of a model hydrophobic solute. We study the solvation and salting-in of benzene in urea-water mixtures where the interactions will be described by means of single-site coarse-grained models. Studies of solutes in the mixtures of urea and water are chosen in this work because urea is a well-known chemical denaturant for proteins[20–23] and preferentially interacts with hydrophobic groups[24,25] and peptide backbones.[23,26,27] Biomacromolecules in water can be unfolded upon addition of urea because of the preferential binding of urea to the polar and the non-polar groups which favors solvent exposure of these groups and decreases the solvation free-energy of the macromolecules in water. In this paper the salting-in of benzene in urea-water solutions will be studied at different urea concentrations (from 6 m to 12 m) at infinite solute dilution and at two finite solute concentrations for benzene in 4-8 m urea-water solution. We use the binary urea/water CG solvent model derived previously[28] and newly parameterize solute-solvent, solute-cosolvent and solute-solute interactions. We also report the free-energy associated with the growth of the benzene clusters in urea-water mixtures with all-atom and CG models. In addition to that we also study the representability of the structure-based coarse-grained models in terms of pressure and isothermal compressibility and the respective convergence of the thermodynamic quantities such as total potential energy for the systems of pure water, binary mixtures of urea and water and ternary mixtures of benzene in urea-water at infinite benzene dilution.

In a previous study,[28] single-site models of urea in water have been developed where the coarse-grained models were able to reproduce (with respect to the atomistic simulations) the pair correlations between solution components as well as the derivatives of the solvation free energies of urea in water with varying urea concentration. In that paper, the authors proposed a new method for coarse-graining, called KB-IBI (Kirkwood-Buff Iterative Boltzmann Inversion), which uses the reference (atomistic) RDFs and the integrals of the RDFs over volume, called Kirkwood-Buff integrals (KBIs), as target properties to develop coarse-grained potentials iteratively. The method can be understood as an extension of IBI and uses KBIs as additional target quantities in the derivation of the coarse-grained models. The rationale behind this method lies in the Kirkwood-Buff theory,[29] which relates KBIs to the macroscopic thermodynamic quantities like isothermal compressibility, partial molar volumes and derivatives of chemical potentials or activity coefficients with composition through KBIs (G_{ij}) defined as

$$\mathrm{G_{ij}} = 4\pi \int_0^\infty \left[\mathrm{g_{ij}}(r) - 1\right] r^2 dr, \tag{5.1}$$

where $\mathrm{g_{ij}}(r)$ is the RDF and $\mathrm{G_{ij}}$ the KBI between the particle types i and j present in the system. The KBI can physically be interpreted as the affinity between the particle types i and j and $\rho_j \mathrm{G_{ij}}$ gives the excess coordination number of particle type j around particle type i where ρ_j is the particle number density of particle type j. In a binary system of solvent (water(w)) and cosolvent (urea(u)) the variation in the molar activity coefficient (γ_u) of the cosolvent at constant pressure p and temperature T is given by[30]

$$\left(\frac{\partial \ln \gamma_\mathrm{u}}{\partial \ln \rho_\mathrm{u}}\right)_{p,T} = -\frac{\rho_\mathrm{u}\left(\mathrm{G_{uu}} - \mathrm{G_{uw}}\right)}{1 + \rho_\mathrm{u}\left(\mathrm{G_{uu}} - \mathrm{G_{uw}}\right)}, \tag{5.2}$$

where ρ_u is the cosolvent number density. Similarly, in a ternary system with solute (b), solvent (w) and cosolvent (u), the KBIs can be related to the derivative of the

solvation free energy (ΔG_{b}) of the solute with respect to the change in the mole-fraction of the cosolvent (x_{u}) at infinite solute dilution,

$$\lim_{\rho_{\mathrm{b}}\to 0}\left(\frac{\partial \Delta G_{\mathrm{b}}}{\partial x_{\mathrm{u}}}\right)_{p,T} = \frac{RT\left(\rho_{\mathrm{w}}+\rho_{\mathrm{u}}\right)^2}{\eta}\left(\mathrm{G_{bw}}-\mathrm{G_{bu}}\right), \tag{5.3}$$

where R is the gas constant, $\eta = \rho_{\mathrm{w}} + \rho_{\mathrm{u}} + \rho_{\mathrm{w}}\rho_{\mathrm{u}}\left(G_{\mathrm{ww}} + G_{\mathrm{uu}} - 2G_{\mathrm{uw}}\right) > 0$ for stable solutions, and "ρ"-s are the number densities of the components present in the system. If the cosolvent affinity of the solute ($\mathrm{G_{bu}}$) exceeds its water affinity ($\mathrm{G_{bw}}$), in other words, if $\mathrm{G_{bu}} > \mathrm{G_{bw}}$, then $\left(\partial \Delta G_{\mathrm{b}}/\partial x_{\mathrm{u}}\right)_{p,T} < 0$, i.e. the solvation free energy of the solute decreases with increasing cosolvent concentration. This phenomenon is called "salting-in". The opposite phenomenon where $\mathrm{G_{bu}} < \mathrm{G_{bw}}$, is referred to as "salting-out" and corresponds to preferential hydration of the solute. Benzene is salted-in by urea in water.[31] From Eq. 5.2 and Eq. 5.3 it is clear that to study the salting-in processes of the solute with coarse-grained descriptions we need to have coarse-grained models which reproduce the respective KBIs. The KB-IBI method has been introduced to provide coarse-grained potentials that reproduce RDFs and KBIs in mixtures and is therefore particularly useful in this context.

5.3 Computational Details

All-atom systems were simulated using GROMACS molecular dynamics (MD) package.[32] For urea a Kirkwood-Buff derived force-field was used.[33] Gromos43a1 parameters[34] for benzene were used. Water was simulated with the SPC/E potential.[36] NpT simulations were performed at a temperature of 300 K and a pressure of 1 bar. The temperature was controlled using a Nose-Hoover thermostat[37,38] with a relaxation time of 0.5 ps. The pressure was kept constant using the Parrinello-Rahman barostat[39] with a coupling time of 3 ps. The particle mesh Ewald (PME)[40] method was used for calculating the electrostatic interactions. The cut-off radius for all non-bonded interactions was set at 1 nm. The equations of motion were integrated using a leap-frog integrator with a 2 fs time-step and 100 ns long trajectories were obtained for all the binary and ternary mixtures (5 ns long trajectory for pure water). For all the simulations of the binary and ternary mixtures cubic boxes with periodic boundary conditions containing 11111 water molecules were studied, the number of the benzene and urea molecules was varied accordingly

(details of the numbers of the molecules are listed in Table 5.1). Simulation of pure water was performed using 2180 water molecules.

For coarse-grained MD simulations, a leap-frog stochastic algorithm was used with an inverse friction constant of 0.2 ps. Coarse-grained systems were simulated under constant NVT conditions (300 K) where the average volumes of the corresponding atomistic simulations were used. The time step for integration was 4 fs and 20 ns long trajectories were accumulated. The cut-off for non-bonded interactions was set to 1.4 nm. Parameterization of the coarse-grained force-fields, both for IBI and KB-IBI, was done using VOTCA 1.2 package.[41] For the ternary mixtures of benzene, urea and water $40 - 60$ KB-IBI iterations were preceded by 30 iterations of IBI (with 10 ns long simulations per iteration for IBI and KB-IBI). Pressure and KB-IBI ramp corrections for the binary urea-water mixture were performed with 2 ns long trajectories per iteration while for pure water all the parameterizations were done with 0.5 ns trajectories per iteration. The pre-factors of the KB-IBI ramp-potentials[28] were chosen to be $\approx$ 0.5 kJ nm^{-3} mol^{-1} for the potentials involving benzene, $\approx$ 0.4 kJ nm^{-3} mol^{-1} for the potentials involving urea in the binary urea-water systems and $\approx$ 10.0 kJ nm^{-3} mol^{-1} for the systems with pure water.

Atomistic and coarse-grained KBIs were calculated by taking the average of the running KBIs ($\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$, r being the radial distance) between r =1.0 and 1.4 nm.

5.4 Results and Discussion

5.4.1 Salting-in at infinite solute dilution

To study salting-in of the solute at infinite dilution in urea-water solutions with coarse-grained single-site models we have run both atomistic and coarse-grained simulations of urea-water using simulation boxes with varying urea concentrations containing just 1 solute (benzene) molecule. For all the coarse-grained systems, the solvent models (effective pair potentials for urea-urea, urea-water and water-water interactions at different urea concentrations) were taken from the simulations of binary urea-water solutions,[28] while the solute-solvent and solute-cosolvent potentials (solute-urea and solute-water) were newly obtained. Systems with a single benzene molecule in urea-water solution were studied with four different urea concentrations, namely 6, 8, 10 and 12 m (molality). For the simulations of 6 m and 8

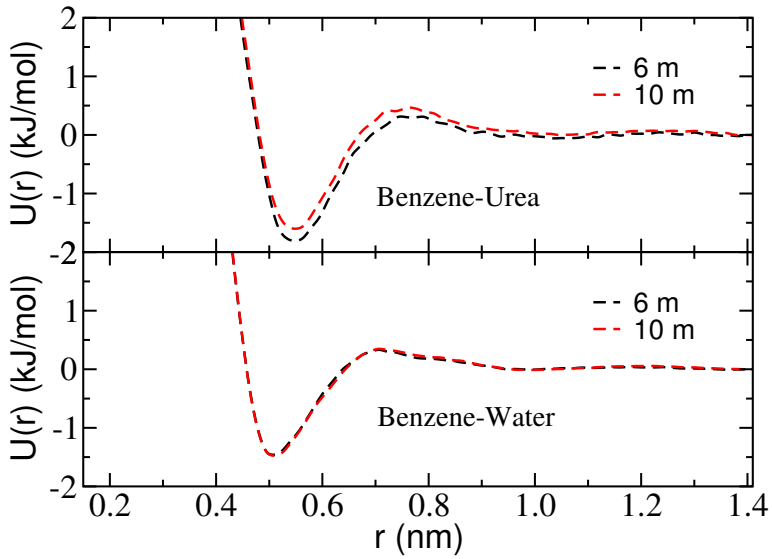


Figure 5.1: Single-site KB-IBI solute-solvent potentials used in coarse-grained molecular dynamics simulations of 6-12 m aqueous urea solutions containing one benzene molecule.

m urea we used the CG solvent model previously derived for pure urea-water solutions at 6 m urea concentration.[28] For the systems with 10 m and 12 m urea, the CG solvent model[28] derived at 10 m urea concentration was used. This choice reflects the concentration transferability of the CG urea/water model which is limited to approximately 2 m concentration windows. For all concentrations of urea, the solute-solvent interactions (benzene-urea and benzene-water) were parameterized using KB-IBI. The solute-solvent KB-IBI potentials obtained at 6 m and 10 m urea are shown in Fig. 5.1. All the potentials parameterized at different urea concentrations are shown in Fig. 5.2.

For all the concentrations of urea, the RDFs and KBIs between benzene and urea and between benzene and water were reproduced in agreement with the atomistic model. Fig. 5.3 shows the RDF between benzene and urea at 6 m urea concentration, both from atomistic and coarse-grained simulations, and the corresponding running KBIs as a function of distance. Similarly, Fig. 5.4 shows the RDF and the corresponding running KBIs between benzene and water at 6 m urea concentration. The KBI between benzene and urea exceeds the KBI between benzene and water, which indicates stronger affinity between benzene and urea over that between benzene and water. So using Eq. 5.3 it confirms the fact that benzene is salted-in by urea in water.

The solvent/cosolvent RDFs were also well reproduced in agreement with the atomistic model though the solvent/cosolvent models were not re-parameterized

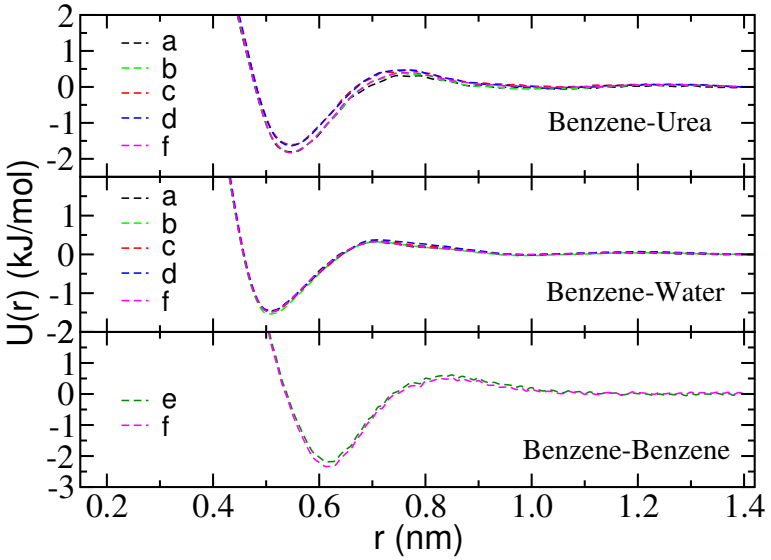


Figure 5.2: Single-site KB-IBI solute-solvent potentials used in coarse-grained molecular dynamics simulations of 6-12 m aqueous urea solutions with single benzene molecule and with finite concentration of benzene. Legends denote different coarse-graining schemes as: **a**– Benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI. **b**– Benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 8 m urea-water using KB-IBI. **c**– Benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 10 m urea-water using KB-IBI. **d**– Benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 12 m urea-water using KB-IBI. **e**– Benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI; benzene-benzene potential is obtained from 0.25 m benzene in 6 m urea-water solution. **f**– Benzene-urea, benzene-water and benzene-benzene potentials are obtained from 0.25 m benzene in 6 m urea-water solution. For the schemes **a**, **b**, **e** and **f** urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water and for the schemes **c** and **d** urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 10 m urea in water.

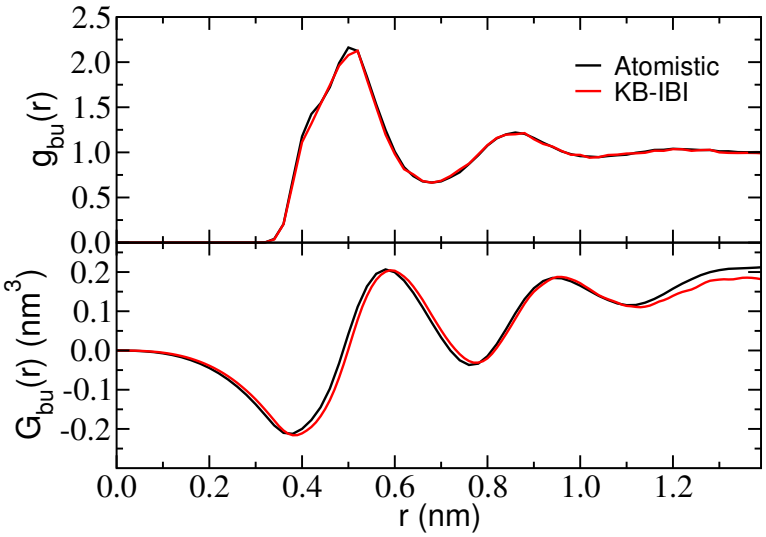


Figure 5.3: Radial distribution functions $\mathrm{g_{bu}}(r)$ and running KB integrals $\mathrm{G_{bu}}(r) = 4\pi \int_0^r [\mathrm{g_{bu}}(s) - 1]\, s^2 ds$ between benzene and urea for a single benzene molecule in 6 m urea-water solution. Urea-urea, urea-water and water-water potentials were obtained by KB-IBI from an atomistic simulation trajectory of 6 m urea in water[28]; benzene-urea and benzene-water potentials were newly parameterized using KB-IBI from simulations of a single benzene molecule in 6 m urea-water solution.

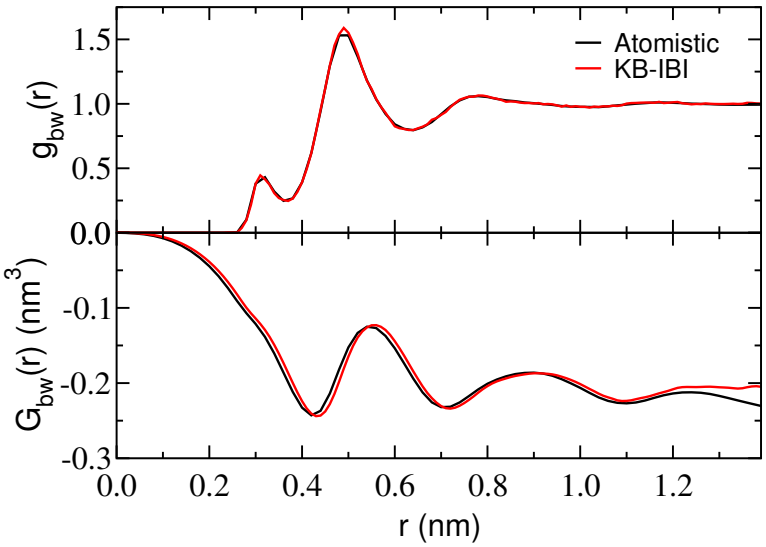


Figure 5.4: Radial distribution functions $\mathrm{g_{bw}}(r)$ and running KB integrals $\mathrm{G_{bw}}(r) = 4\pi \int_0^r [\mathrm{g_{bw}}(s) - 1]\, s^2 ds$ between benzene and water for a single benzene molecule in 6 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI from an atomistic simulation trajectory of 6 m urea in water[28]; benzene-urea and benzene-water potentials were newly parameterized using KB-IBI from simulations of a single benzene molecule in 6 m urea-water solution.

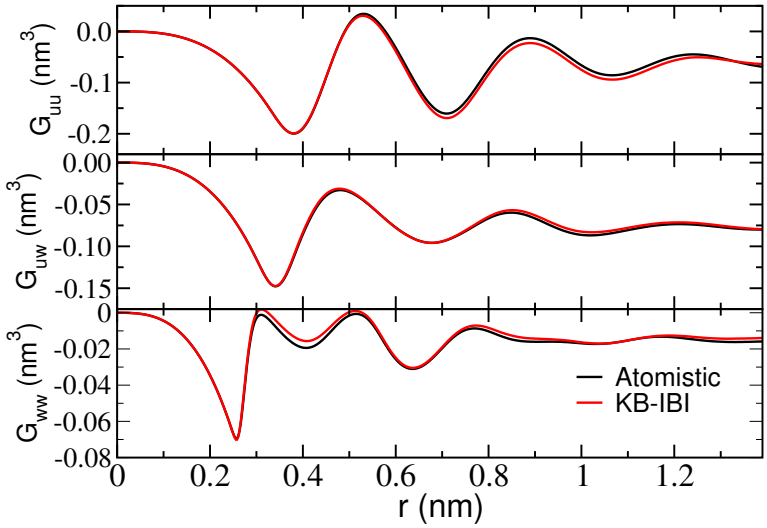


Figure 5.5: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between urea-urea (uu), urea-water (uw) and water-water (ww) for a single benzene molecule in 6 m urea-water solution. Urea-urea, urea-water and water-water potentials were obtained by KB-IBI method from binary mixture of 6 m urea in water[28]; benzene-urea and benzene-water potentials were newly parameterized from single benzene molecule in 6 m urea-water using KB-IBI.

after the insertion of the single solute molecule. Fig. 5.5 shows the KBIs between urea-urea, urea-water and water-water for the system of 1 benzene in 6 m urea-water solution, both atomistic and coarse-grained. Good agreement is found, as expected, because the solvent/cosolvent RDFs and KBIs do not deviate significantly from those of a pure urea-water solution at 6 m urea concentration.[28] The derivative of the solvation free-energy of benzene with respect to the urea concentration at 6 m urea was calculated using Eq. 5.3. The value of $(\partial \Delta G_\mathrm{b}/\partial x_\mathrm{u})_{p,T}$ obtained from the coarse-grained simulation with benzene ($-22.1\ \mathrm{kJ/mol}$) agrees well with the result from atomistic simulation ($-23.4\ \mathrm{kJ/mol}$).

For the simulations of 1 benzene molecule in 8 m urea-water we used urea-urea, urea-water and water-water CG potentials derived for a binary 6 m urea-water solution (as mentioned earlier) and parameterized benzene-urea and benzene-water potentials using KB-IBI. Fig. 5.6 shows that the benzene-urea and benzene-water KBIs are reproduced and the solvent-solvent KBIs are also in satisfactory agreement with the atomistic simulation results (Fig. 5.7). The derivative of the solvation free-energy of benzene with urea concentration at 8 m urea (all-atom: $-9.3\ \mathrm{kJ/mol}$) is also closely reproduced (coarse-grained: $-9.9\ \mathrm{kJ/mol}$). Further, benzene-urea and benzene-water potentials were parameterized at 10 m and 12 m urea concentration

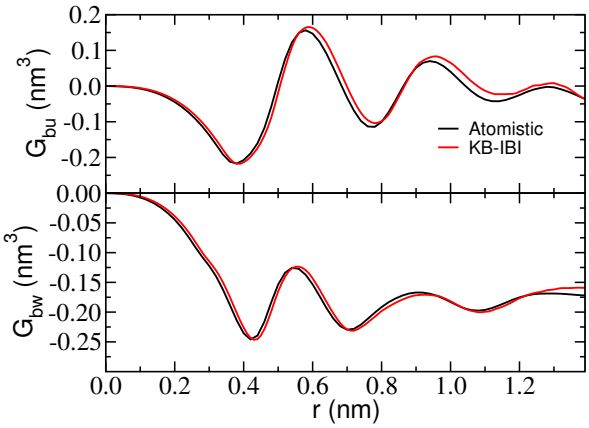


Figure 5.6: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between benzene-urea (bu) and benzene-water (bw) for a single benzene molecule in 8 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 8 m urea-water using KB-IBI.

using a fixed solvent model derived for the binary 10 m urea-water solution and the representability of the model was validated in terms of reproducing the KBIs and the variation of solvation free energy of benzene (see Fig. 5.8 and Fig. 5.9). The derivatives of the benzene solvation free-energies with urea concentration are listed in Table 5.1.

5.4.2 Salting-in of benzene at finite concentrations

To study salting-in of solutes at finite solute concentrations the additional interaction that comes into play is the solute-solute interaction. To find the solute-solute interaction we followed two procedures: 1) taking the solvent/cosolvent interactions from the binary urea-water system while using the solute-solvent/cosolvent interactions from the systems of one solute in urea-water (described in the previous section) and using KB-IBI to find the solute-solute interaction or 2) taking the solvent/cosolvent interactions from the binary urea-water system and using KB-IBI to parameterize solute-solute, solute-solvent and solute-cosolvent interactions. To test both procedures we studied a system of 0.25 m benzene in 6 m urea-water (50 benzene molecules, 11111 water and 1200 urea molecules). Fig. 5.10 shows the KBIs between benzene-urea, benzene-benzene and benzene-water obtained using the first method. As we only iteratively updated the benzene-benzene potential using KB-IBI, while keeping all other interactions unaltered, the KBI between benzene-

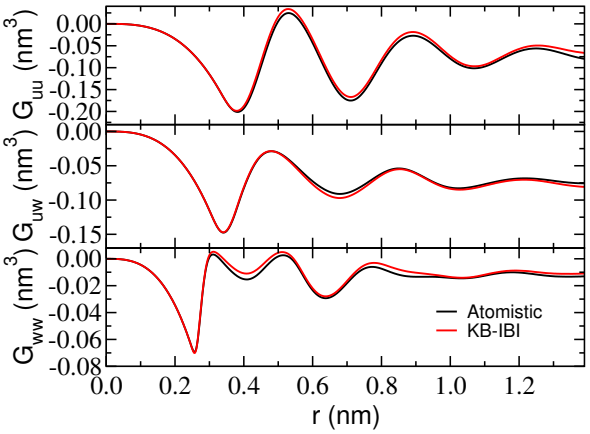


Figure 5.7: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between urea-urea (uu), urea-water (uw) and water-water (ww) for a single benzene molecule in 8 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 8 m urea-water using KB-IBI.

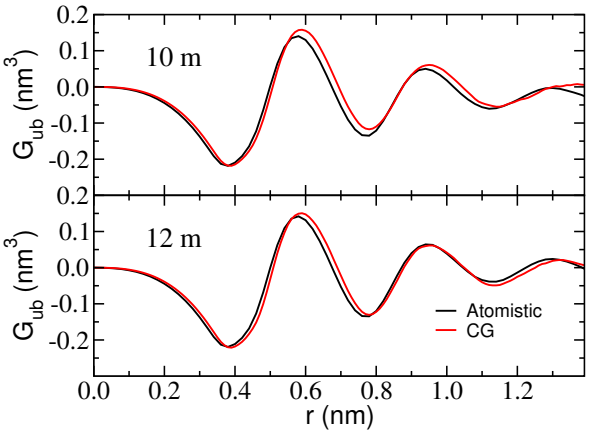


Figure 5.8: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between benzene-urea (bu) for a single benzene molecule in 10 m and 12 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 10 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 10 m and 12 m urea-water respectively using KB-IBI.

System	N_u	N_b	N_w	G_{bw} (nm^3)	G_{bu} (nm^3)	G_{bb} (nm^3)	G_{uu} (nm^3)	G_{uw} (nm^3)	G_{ww} (nm^3)	$\left(\frac{\partial \Delta G_b}{\partial x_u}\right)_{p,T,\rho_b \to 0}$ (kJ/mol)	$\left(\frac{\partial \mu_b}{\partial N_u}\right)_{p,T}$ (kJ/mol)
$6-$ dil (AA)	1200	1	11111	-0.221	0.168	–	-0.064	-0.079	-0.015	-23.4	–
$6-$ dil (CG)[a]	1200	1	11111	-0.221	0.151	–	-0.069	-0.076	-0.014	-22.1	–
$6-$ dil (CG)[f]	1200	1	11111	-0.208	0.130	–	-0.073	-0.075	-0.015	-21.1	–
$6-$ dil (CG)[g]	1200	1	11111	-0.220	0.182	–	-0.056	-0.084	-0.005	-22.8	–
$8-$ dil (AA)	1600	1	11111	-0.180	-0.021	–	-0.077	-0.074	-0.012	-9.3	–
$8-$ dil (CG)[b]	1600	1	11111	-0.178	-0.003	–	-0.070	-0.077	-0.011	-9.9	–
$10-$ dil (AA)	2000	1	11111	-0.190	-0.026	–	-0.090	-0.069	-0.011	-9.8	–
$10-$ dil (CG)[c]	2000	1	11111	-0.186	-0.018	–	-0.086	-0.070	-0.010	-9.8	–
$12-$ dil (AA)	2400	1	11111	-0.214	-0.002	–	-0.099	-0.066	-0.010	-12.7	–
$12-$ dil (CG)[d]	2400	1	11111	-0.210	-0.008	–	-0.084	-0.070	-0.006	-10.7	–
$4-0.25$ (AA)	800	50	11111	-0.199	0.137	0.407	-0.027	-0.088	-0.015	–	-0.00193
$4-0.25$ (CG)[e]	800	50	11111	-0.221	0.275	0.668	-0.067	-0.084	-0.014	–	-0.00280
$4-0.25$ (CG)[f]	800	50	11111	-0.205	0.150	0.880	-0.066	-0.070	-0.015	–	-0.00207
$6-0.25$ (AA)	1200	50	11111	-0.232	0.139	0.759	-0.062	-0.082	-0.010	–	-0.00187
$6-0.25$ (CG)[e]	1200	50	11111	-0.253	0.234	0.751	-0.066	-0.083	-0.009	–	-0.00239
$6-0.25$ (CG)[f]	1200	50	11111	-0.223	0.137	0.738	-0.065	-0.082	-0.010	–	-0.00184
$8-0.25$ (AA)	1600	50	11111	-0.228	0.076	0.460	-0.081	-0.075	-0.008	–	-0.00155
$8-0.25$ (CG)[e]	1600	50	11111	-0.280	0.236	0.282	-0.062	-0.084	-0.003	–	-0.00230
$8-0.25$ (CG)[f]	1600	50	11111	-0.270	0.194	0.793	-0.063	-0.084	-0.004	–	-0.00199
$4-0.50$ (AA)	800	100	11111	-0.237	0.187	0.793	-0.040	-0.092	-0.009	–	-0.00205
$4-0.50$ (CG)[f]	800	100	11111	-0.243	0.154	1.074	-0.076	-0.084	-0.009	–	-0.00195
$6-0.50$ (AA)	1200	100	11111	-0.247	0.116	0.775	-0.064	-0.087	-0.004	–	-0.00165
$6-0.50$ (CG)[e]	1200	100	11111	-0.286	0.275	0.875	-0.061	-0.092	-0.001	–	-0.00240
$6-0.50$ (CG)[f]	1200	100	11111	-0.243	0.174	0.601	-0.066	-0.086	-0.004	–	-0.00196
$8-0.50$ (AA)	1600	100	11111	-0.254	0.069	0.857	-0.085	-0.076	-0.002	–	-0.00142
$8-0.50$ (CG)[f]	1600	100	11111	-0.263	0.131	0.749	-0.065	-0.085	0.001	–	-0.00160

Table 5.1: Comparison between the atomistic and coarse-grained (KB-IBI) force-fields in terms of the Kirkwood-Buff integrals (KBIs) and the derivatives of the solvation free-energy or chemical potential of benzene with varying urea concentrations. Subscripts b, u, and w stand for benzene, urea and water respectively. G_{ij}-s are the KBIs and N_i-s are the numbers of the molecules of different species. $\left(\frac{\partial \Delta G_b}{\partial x_u}\right)_{p,T,\rho_b \to 0}$ is the derivative of the solvation free-energy of benzene with varying urea mole-fraction at infinite dilution of benzene and $\left(\frac{\partial \mu_b}{\partial N_u}\right)_{p,T}$ is the derivative of the chemical potential of benzene with number of urea molecules at finite concentrations of benzene. The labels for the systems denote the molal concentrations of urea and benzene respectively ($6-0.25$ being 0.25 molal benzene in 6 m urea-water solution; dil: infinite dilution of benzene). AA: all-atom; CG: coarse-grained. Superscripts **a** to **f** refer to the different KB-IBI coarse-graining schemes as: [a] Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI. [b] Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 8 m urea-water using KB-IBI. [c] Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 10 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 10 m urea-water using KB-IBI. [d] Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 10 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 12 m urea-water using KB-IBI. [e] Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI; benzene-benzene potential is obtained from 0.25 m benzene in 6 m urea-water solution. [f] Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea, benzene-water and benzene-benzene potentials are obtained from 0.25 m benzene in 6 m urea-water solution. [g] Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water with pressure corrections applied to water-water potential; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI.

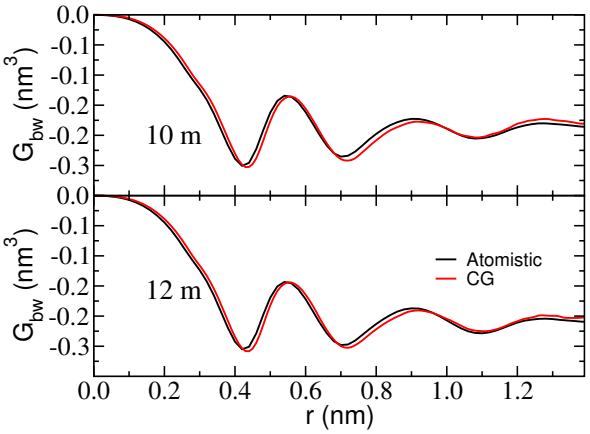


Figure 5.9: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between benzene-water (bw) for a single benzene molecule in 10 m and 12 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 10 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 10 m and 12 m urea-water respectively using KB-IBI.

benzene was reproduced in agreement with the all-atom result, but benzene-urea and benzene-water running KBIs could only be reproduced at distances smaller than 0.8 nm while discrepancies were observed in the larger distance region. This is due to the fact that the atomistic KBIs between benzene-urea and benzene-water are significantly different at infinite benzene dilution and at finite concentration of benzene. The KB-IBI solute-solvent potentials thus show limited transferability. The solvent/cosolvent RDFs and KBIs were however reproduced, as shown in Fig. 5.11, indicating that the fixed KB-IBI solvent model is transferable to finite solute concentrations. For finite concentrations of benzene we calculated the derivative of the benzene chemical potential (μ_b) with respect to the change in the number of urea molecules ($\mathrm{N_u}$) using[30]

$$\mu_{\mathrm{bu}} = \left(\frac{\partial \mu_\mathrm{b}}{\partial \mathrm{N_u}}\right)_{\mathrm{N_b},\mathrm{N_w},p,T} = \frac{k_B T[1 + \rho_w(\mathrm{G_{bu}} + \mathrm{G_{ww}} - \mathrm{G_{bw}} - \mathrm{G_{uw}})]}{V\eta^{finite}}, \tag{5.4}$$

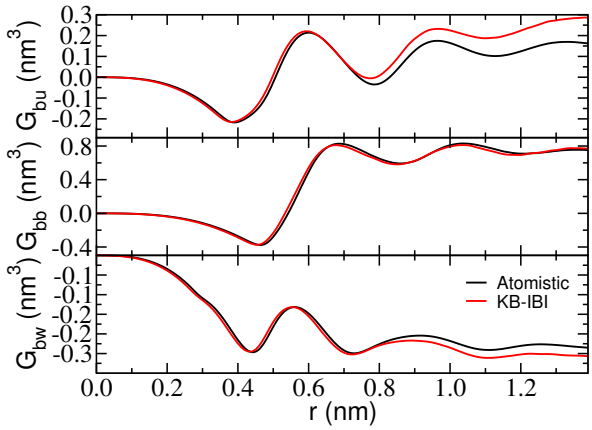


Figure 5.10: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between benzene-urea (bu), benzene-benzene (bb) and benzene-water (bw) for 0.25 m benzene in 6 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI; benzene-benzene potential is obtained from 0.25 m benzene in 6 m urea-water solution.

where k_B is Boltzmann constant and η^{finite} is given by

$$\begin{aligned}\eta^{finite} =& \rho_{\mathrm{b}} + \rho_{\mathrm{u}} + \rho_{\mathrm{w}} + \rho_{\mathrm{b}}\rho_{\mathrm{u}}\Delta_{\mathrm{bu}} + \rho_{\mathrm{u}}\rho_{\mathrm{w}}\Delta_{\mathrm{uw}} + \rho_{\mathrm{b}}\rho_{\mathrm{w}}\Delta_{\mathrm{bw}} \\ & - \frac{1}{4}\rho_{\mathrm{b}}\rho_{\mathrm{u}}\rho_{\mathrm{w}}\left(\Delta_{\mathrm{bu}}^2 + \Delta_{\mathrm{uw}}^2 + \Delta_{\mathrm{ww}}^2 - 2\Delta_{\mathrm{bw}}\Delta_{\mathrm{uw}} - 2\Delta_{\mathrm{bu}}\Delta_{\mathrm{bw}} - 2\Delta_{\mathrm{bu}}\Delta_{\mathrm{uw}}\right)\end{aligned} \tag{5.5}$$

with $\Delta_{\mathrm{ij}} = \mathrm{G_{ii}} + \mathrm{G_{jj}} - 2\mathrm{G_{ij}}$ and V being the volume of the system. μ_{bu} for the CG system was $-0.00239\ \mathrm{kJ/mol}$ which compares with the atomistic simulation result of $-0.00187\ \mathrm{kJ/mol}$.

To test the transferability of the model with different urea and benzene concentrations we have applied these coarse-grained potentials to 4 m and 8 m urea-water solutions (with 0.25 m benzene) and to 6 m urea-water solution with 0.5 m benzene. For all the systems benzene-urea KBIs were significantly over-estimated and benzene-water KBIs were slightly under-estimated. The solute-solute KBIs also showed deviations from all-atom simulations. The solvent KBIs (urea-urea, urea-water and water-water) were reproduced reasonably except for the urea-urea KBI at 4 m urea concentration which shows the limitation in the transferability of the coarse-grained binary urea-water model. The corresponding KBIs and the deriva-

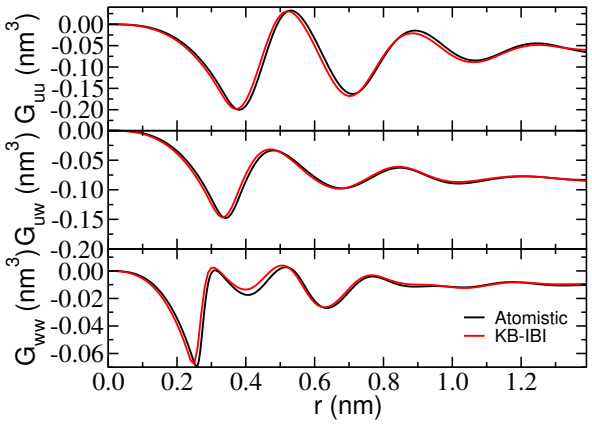


Figure 5.11: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between urea-urea (uu), urea-water (uw) and water-water (ww) for a 6 m urea-water solution with 0.25 m benzene. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water; benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI; benzene-benzene potential is obtained from 0.25 m benzene in 6 m urea-water solution.

tives of the chemical potential of benzene with varying solution composition are listed in Table 5.1 (see superscript e).

The second method where we updated all three potentials, namely between benzene-urea, benzene-water and benzene-benzene, while again keeping the solvent/cosolvent model fixed, yielded a significantly better match with the atomistic KBIs, as expected. The results are shown in Fig. 5.12 where all three KBIs between benzene-urea, benzene-benzene and benzene-water are well matching to the atomistic ones. The urea-urea, urea-water and water-water KBIs were also in agreement with their atomistic values (data not shown). The derivative of the chemical potential of benzene with varying urea number at 6 m urea concentration, μ_{bu}, was $-0.00184\ \mathrm{kJ/mol}$ which was very close to the atomistic value $-0.00187\ \mathrm{kJ/mol}$.

Further we applied these potentials obtained from the simulations of 0.25 m benzene in 6 m urea-water to systems of 0.25 m benzene in 4, 6, and 8 m urea-water and also to systems of 0.5 m benzene in 4, 6, and 8 m urea-water to test the transferability of our potentials with different solute and cosolvent concentrations. The coarse-grained benzene-urea KBIs were found to be slightly off in comparison to their atomistic counterparts but benzene-water, urea-urea (except for 4 m urea concentrations), urea-water and water-water KBIs were well reproduced. The benzene-benzene KBIs suffer from poorer statistics (for both atomistic and coarse-grained

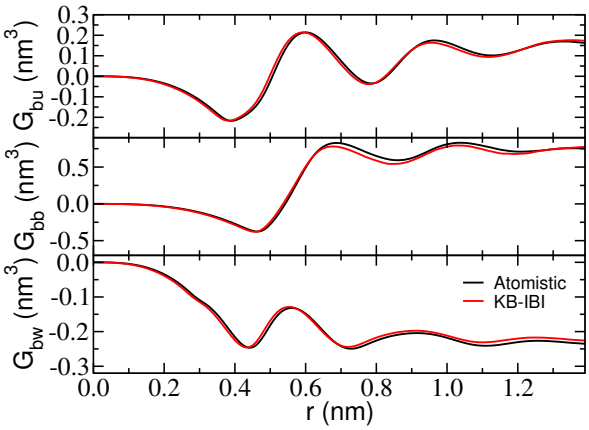


Figure 5.12: Running KB integrals $G_{ij}(r) = 4\pi \int_0^r [g_{ij}(s) - 1]\, s^2 ds$ between benzene-urea (bu), benzene-benzene (bb) and benzene-water (bw) for 0.25 m benzene in 6 m urea-water solution. Here all three potentials between benzene-urea, benzene-benzene and benzene-water are updated using KB-IBI at 0.25 m benzene in 6 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water.

simulations) and were not for all the systems well reproduced by the coarse-grained potentials, although the mismatches in the benzene-benzene KBIs do not affect the results significantly while calculating the thermodynamic quantities like the derivative of the chemical potential of benzene as the concentration of benzene is very low (see Eq. 5.4). The derivatives of the chemical potential of benzene were closely reproduced for almost all the systems when compared to the atomistic values (results listed in Table 5.1, see superscript **f**). For 0.5 m benzene in 6 m urea-water the KBIs between benzene-urea, benzene-water and benzene-benzene are shown in Fig. 5.13. To verify the transferability of the KB-IBI potentials obtained at a finite concentration of solute to the systems with infinite dilution of solute, we also applied this set of potentials to a system of 6 m urea-water containing a single molecule of benzene and we found benzene-urea and benzene-water KBIs were in reasonable agreement with all-atom results. This result is particularly interesting as benzene-urea and benzene-water coarse-grained potentials which were parameterized at infinite dilution of benzene showed relatively poorer transferability when applied to a system of finite benzene concentration but the solute-solvent potentials obtained at a finite concentration of benzene yield much better transferability to a system with infinite solute-dilution.

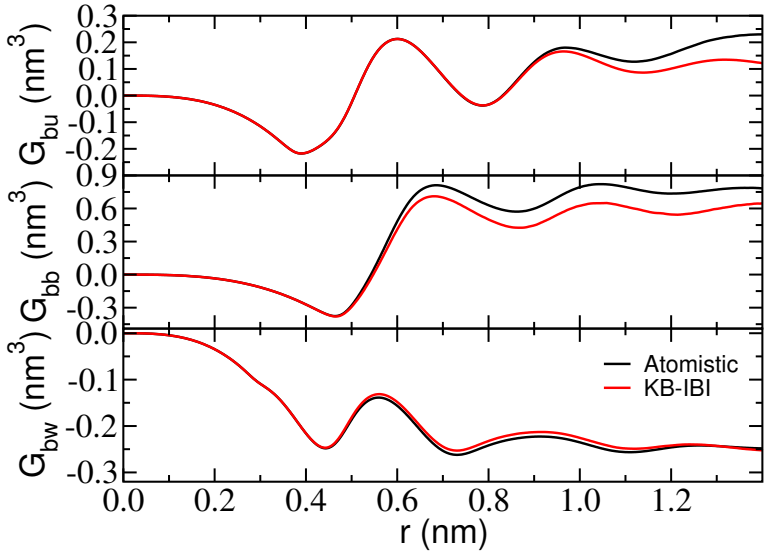


Figure 5.13: Running KB integrals $\mathrm{G_{ij}}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ between benzene-urea (bu), benzene-benzene (bb) and benzene-water (bw) for 0.5 m benzene in 6 m urea-water solution. Here all three potentials between benzene-urea, benzene-benzene and benzene-water are updated using KB-IBI at 0.25 m benzene in 6 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water.

5.4.3 Cluster analysis of benzene in urea-water solution

Benzene, at low concentration in water, forms hydrophobic clusters.[42] Addition of urea to benzene-water systems disfavors the formation of clusters as urea preferentially solvates the benzene solutes thereby reducing the hydrophobic aggregation of benzene in the solution.[43] In this section, we examine the representability of the CG model and compare clustering data with results obtained from all-atom simulations. The free-energy of benzene cluster growth can be defined as

$$\Delta \mathrm{G_{growth}} = -RT \ln \frac{[n_{S+1}]C_0}{[n_S][n_1]} \tag{5.6}$$

where $[n_S]$ denotes the average equilibrium concentration of clusters of S benzene molecules with C_0 being the unit of the concentration. If urea disfavors clustering of benzene in water, then one can expect $\Delta \mathrm{G_{growth}}$ to increase with increasing urea concentrations. A cluster analysis was carried out at a higher benzene concentration (0.5 m) in urea-water with different urea concentrations (4, 6 and 8 m).

The average numbers of clusters of size S per frame were calculated using a distance criterion for the clustered benzene molecules. The cut-off distance to find the benzene molecules belonging to a cluster was chosen to be 0.73 nm which corresponds to the first minimum after the first peak of the benzene-benzene RDFs. The solvent-cosolvent CG model was taken from pure 6 m urea-water solution and the benzene-benzene, benzene-urea and benzene-water potentials were taken from the simulations of 0.25 m benzene in 6 m urea-water by updating benzene-benzene, benzene-urea and benzene-water potentials (mentioned as second method in the previous section and by superscript **f** in Table 5.1). The logarithm of the average numbers of clusters (n_S) with the cluster-size (S) is plotted in Fig. 5.14 for three urea concentrations. We find that the atomistic and CG results are in reasonable agreement with each other though the CG models were not parameterized based on these quantities. For 6 m urea concentration, we see that the results from the CG model slightly deviate from the atomistic ones for the cluster sizes 4 or bigger. But we see an over-all decrease in the number of clusters with increasing urea concentration. Further we calculated the free-energy of cluster growth using Eq. 5.6. The results are plotted in Fig. 5.15 for the atomistic and CG force-field models. For all the urea concentrations, good agreement between the KB-IBI and atomistic results can be observed. When compared with the results obtained from the IBI method (only solute-solute and solute-solvent potentials are obtained from 6 m urea-water solution with 0.25 m benzene using the IBI method; urea-urea, urea-water and water-water potentials are obtained from 6 m urea-water solution using KB-IBI) we find that IBI potentials over-estimate the free-energy of the growth of the benzene clusters for the bigger cluster sizes. The free-energies of the benzene cluster-growth in benzene-water systems are plotted in Fig. 5.16 for 0.2 m and 0.5 m benzene concentrations where the coarse-grained potentials are obtained from KB-IBI of pure benzene-water solutions at 0.2 m benzene concentration. Overall we find an increase in the free-energy of the growth of the benzene clusters (at 0.5 m benzene concentration) with increasing urea concentrations which also shows that urea disfavors the formation of benzene clusters and makes benzene more soluble in water. The relatively large error-bars in the estimation of the free-energies can be explained by two possible causes: 1) the bigger clusters suffer from poor sampling as they occur rarely and/or 2) the sampling time for the atomistic/coarse-grained simulations was not enough.[44]

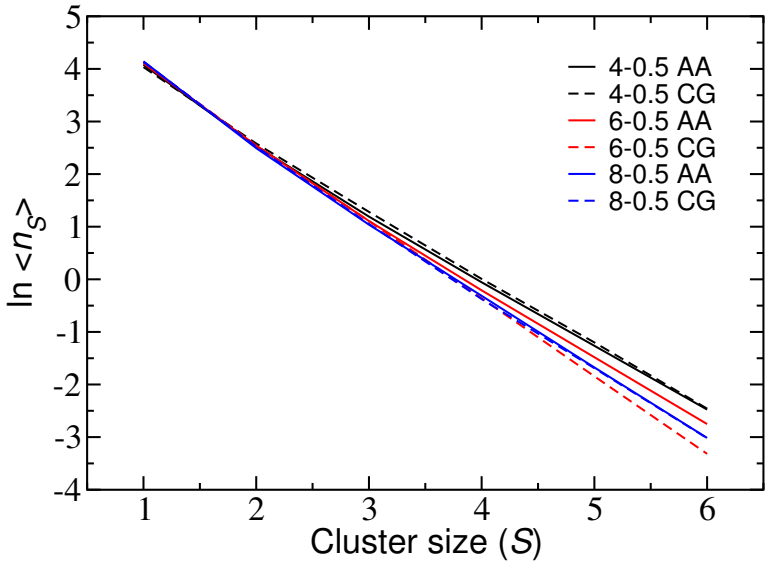


Figure 5.14: Logarithm of the average numbers of the benzene clusters per frame with the size of the cluster in urea-water solutions. AA: all-atom, CG: coarse-grained (KB-IBI, solute-solute and solute-solvent potentials are obtained from 0.25 m benzene in 6 m urea-water solution; see superscript f in Table 5.1). The numbers in the legends represent the concentrations (in molality) of urea and benzene respectively.

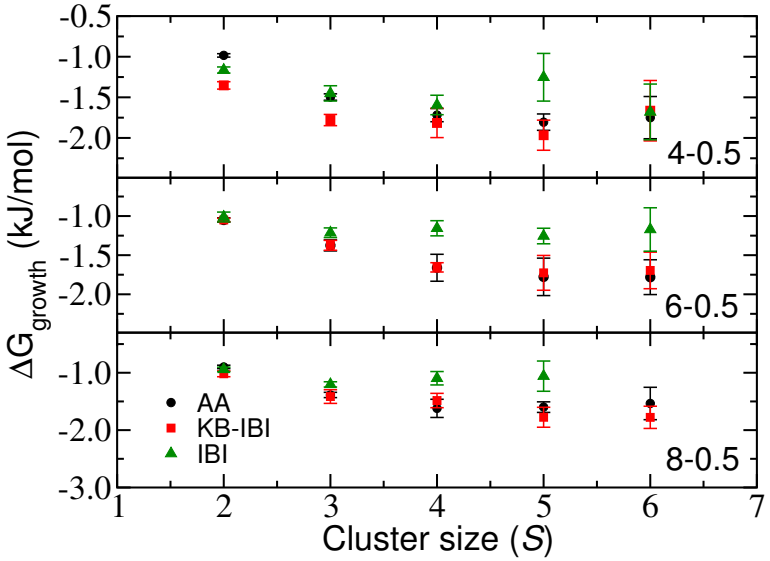


Figure 5.15: Free energy of benzene cluster growth versus cluster size in urea-water solutions. AA: all-atom. The numbers in the labels represent the concentrations (in molality) of urea and benzene, respectively. The coarse-grained potentials are obtained using IBI and KB-IBI where solute-solute and solute-solvent potentials are obtained from 0.25 m benzene in 6 m urea-water solution. The KB-IBI-derived solvent model for a 6 m urea in water solution was used in all calculations.

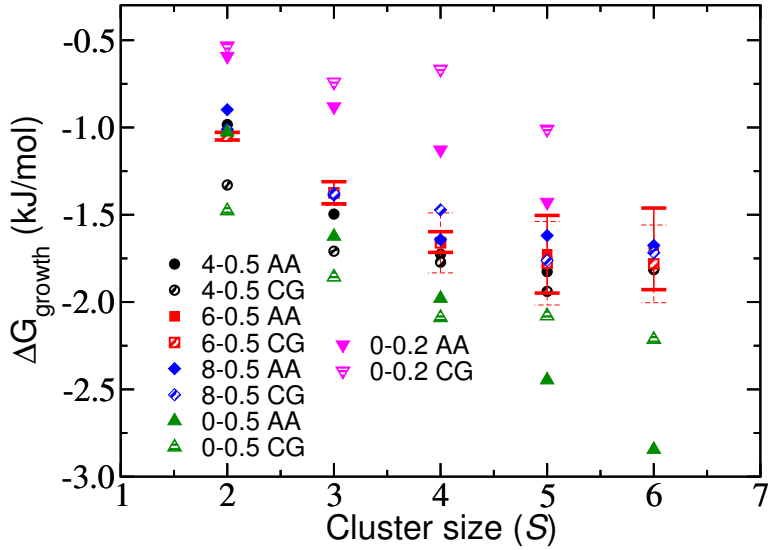


Figure 5.16: Free energy of the growth of the benzene clusters with the size of the cluster in urea-water solutions. AA: all-atom, CG: coarse-grained. The numbers in the legends represent the concentrations (in molality) of urea and benzene respectively. For $4-0.5$, $6-0.5$, and $8-0.5$ systems CG potentials are obtained using KB-IBI where solute-solute and solute-solvent potentials are obtained from 0.25 m benzene in 6 m urea-water solution; see superscript **f** in Fig. 5.2. For pure benzene-water systems ($0-0.2$ and $0-0.5$) KB-IBI models are parameterized at pure benzene-water solution of 0.2 m benzene concentration.

5.4.4 Structure-based coarse-graining with additional thermodynamic targets

In this section we discuss some aspects of applying thermodynamic constraints in the IBI and KB-IBI methods. Pressure and isothermal compressibility are considered as additional targets for pure water, while pressure and KBIs are considered for the mixtures with urea. Further, the potentials obtained from binary urea-water mixtures using KB-IBI and pressure correction are applied to obtain solute-solvent potentials for ternary mixtures of benzene in urea-water at infinite benzene dilution.

5.4.4.1 Pure water

The single-site coarse-grained IBI model for pure water, parameterized based on SPC/E water using the oxygen-oxygen RDF, shows a very high pressure and a lower isothermal compressibility than the atomistic SPC/E model.[18] The isothermal compressibility is related to the water-water Kirkwood-Buff integral as

$$\kappa_T = \frac{(1 + \rho_{\mathrm{w}} \mathrm{G}_{\mathrm{ww}})}{\rho_{\mathrm{w}} k_B T} \tag{5.7}$$

A lower isothermal compressibility of the coarse-grained model shows that the IBI potential cannot reproduce the integral of the water-water RDF (KB integrals) accurately. Fig. 5.17 presents the water-water KBI (lower panel) of pure IBI water along with the total potential energy (upper panel) and pressure (middle panel) versus the number of IBI iterations. It can be observed that the potential energy and the pressure of the system are not converged even after hundreds of IBI iterations. Similar convergence problems of the total potential energy have been reported in the literature.[17] The water-water KBI obtained from the IBI model is approximately 6% lower than the atomistic KBI which is shown by the blue horizontal lines in Fig. 5.17 and 5.18. While the water-water RDF converges rapidly after few IBI iterations, the pair potential still changes significantly to give rise to the difference in the pressure, energy or the KBI which are very sensitive to the tail of the pair-potential. When KB ramp corrections[28] are applied to the potential after 60 normal IBI iterations (with no further iterations being applied to match the RDFs) we find a very fast convergence of the KBI and the pressure (Fig. 5.18, red dots). This procedure – referred to as KB-IBI in the legend of Fig. 5.18 – reproduces the KBI accurately while the pressure is significantly reduced and closer to the pressure of the all-atom system (1

bar). If instead a WJK-type pressure correction[18] is applied to the IBI potentials (Fig. 5.18, orange squares) the pressure could be restored to the pressure of the all-atom system approximately, but the KBI deviates from the target value. With another approach we applied KB ramp corrections after the IBI iterations and when the KBI was converged to the target value we simultaneously applied the WJK-type pressure correction and KB ramp corrections in every iteration step. With this approach the final results (Fig. 5.18, green triangles) show a compromise in simultaneously reproducing the KBI and the pressure. The pressure could be reduced by one order of magnitude while the KBI was reproduced within 2% of the reference value. With the KB-IBI procedure[28] or with WJK pressure corrections the total potential energy of the system converges, unlike with IBI, after very few iterations. However, application of KB ramp corrections together with pressure corrections results in slower convergence of the potential energy but the running average of the potential energy with iterations does not vary much unlike IBI. The corresponding data are plotted in Fig. 5.19.

5.4.5 Binary urea-water mixture

Similar to the system of pure water, IBI or KB-IBI models for binary mixtures of urea-water also show very high pressures of the order of a few thousand bar. These unphysical high pressures may cause the system to behave as a highly compressed fluid (rather than a liquid) and may affect processes which depend on pressure and volume fluctuations. In order to parameterize single-site coarse-grained models for urea-water mixtures which reproduce the atomistic pressure, one can apply pressure-corrections to the coarse-grained potentials in a way similar to the pure water system. But from the studies with pure water we have found that simultaneously reproducing the pressure and the KBI is not straight-forward. So one may expect that for the binary mixtures of urea-water all the three KBIs, namely urea-urea, urea-water and water-water, together with the system pressure cannot be reproduced if the combined KB-IBI and pressure correction procedure is applied to all three potentials. From Eq. 5.2 we find that the variation of the molar activity coefficient of urea with urea concentration does not depend on the water-water KBI. Also one can expect that the pressure is mostly sensitive to the water-water potential as the number of water molecules is much larger than the number of urea molecules. So in an attempt to stay consistent with the solvation thermodynamics of urea in water and to simultaneously repair the pressure we have applied a WJK-type pressure

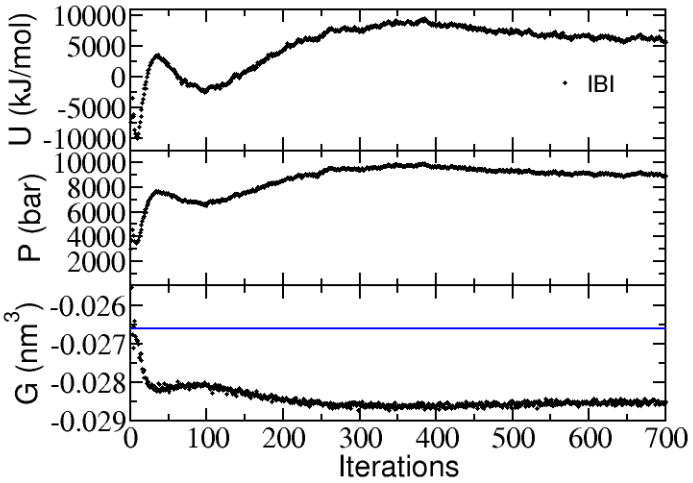


Figure 5.17: Thermodynamic properties of single-site coarse-grained water model obtained from IBI method. Shown are the total pair-potential energy of the system(upper panel), pressure of the system (middle panel) and the Kirkwood-Buff integral between the water molecules (lower panel). The single-site IBI model for pure water shows very slow convergence in the thermodynamic properties with iterations. The pressure amounts to $\approx$ 10000 bar which is much higher than the all-atom simulations (1 bar). Also the water-water Kirkwood-Buff integral, which is related to the isothermal compressibility, does not converge to the all-atom value.

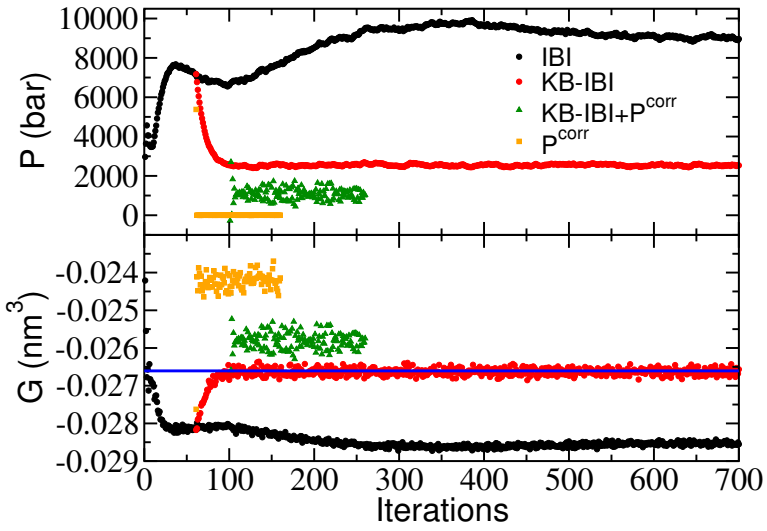


Figure 5.18: System pressure (upper panel) and the water-water Kirkwood-Buff integral (lower panel) for single-site coarse-grained model for water. Black dots are the IBI results (same as in Fig. 5.17). Red dots are the results obtained with the KB-IBI method, applied after 60 IBI iterations; it can be observed that Kirkwood-Buff integrals converge to the all-atomistic value (shown in blue horizontal line) with the KB-IBI method. Orange squares represent the results obtained by using a pressure correction to the IBI potential after 60 normal IBI iterations; the pressure of the system rapidly converges to a value around 1 bar with pressure correction. Green triangles show the data obtained by simultaneously applying KB ramp and pressure corrections after performing 60 normal IBI steps and then 40 normal KB-IBI steps.

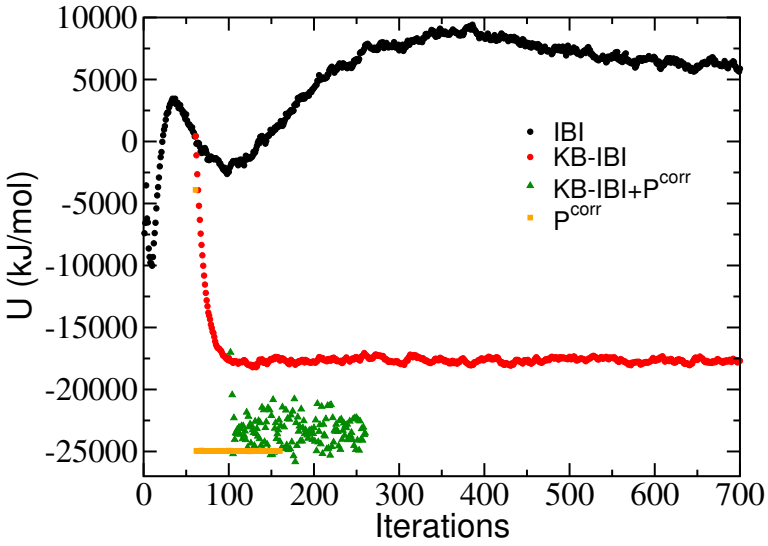


Figure 5.19: Total pair-potential energy of the system of single-site coarse-grained water. Black lines show the KB-IBI potentials without any presure correction to the potentials and the red lines denote the potentials where KB-IBI iterations were performed for urea-urea (uu) and urea-water (uw) potentials and simultaneously pressure correction (WJK type, see main article) was applied to the water-water potentials. All the corrections (KB-IBI and/or pressure) were preceded by 30 normal IBI iterations for all the interactions.

correction to the water-water KB-IBI potential (obtained with KB-IBI of urea-water mixture) without applying any further KB-IBI corrections to it but continuing to perform KB ramp corrections to the urea-urea and urea-water potentials to reproduce urea-urea and urea-water KBIs. With these corrections (pressure-corrected KB-IBI) the system pressure reaches its target value while urea-urea and urea-water KBIs are also reproduced. However, the water-water KBI is not reproduced as expected. The water-water KBI is in fact larger than the all-atom result as the addition of the attractive contributions to the potential in the pressure correction procedure leads to somewhat higher aggregation between the water molecules. The results showing the system pressure and the potential energy versus the number of coarse-grained iterations are plotted in Fig. 5.20 for a system of 6 m urea in water and we find that the system pressure and the total potential energy converge in ≈ 40 pressure-corrected KB-IBI iterations. In Fig. 5.21 we plot the running KBIs between urea-urea, urea-water and water-water obtained with the above procedure. Although the water-water KBI was not accurately reproduced, the derivative of the urea molar activity coefficient with the molar concentration of urea (Eq. 5.2) was reasonably well reproduced ($-0.072\ \mathrm{kJ/mol}$) with respect to the all-atom simulation ($-0.053\ \mathrm{kJ/mol}$).

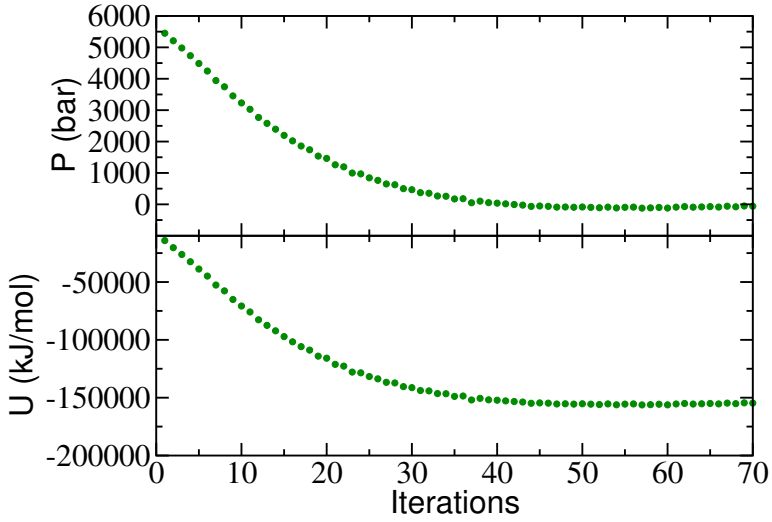


Figure 5.20: System pressure (upper panel) and the total potential energy (lower panel) presented against the number of iterations for a binary mixture of 6 m urea in water. Here we started with KB-IBI urea-urea, urea-water and water-water potentials and applied a pressure correction to the water-water potential without any further KB-IBI correction and continued applying KB ramp corrections to the urea-urea and urea-water potentials.

This compares to standard KB-IBI without pressure correction and standard IBI without pressure correction where -0.055 and $-0.116\ \mathrm{kJ/mol}$ are found, respectively.[28] The newly parameterized urea-urea, urea-water and water-water potentials are plotted in Fig. 5.22 including a comparison with the normal KB-IBI potentials without any pressure correction.[28]

5.4.6 Benzene in urea-water at infinite dilution of benzene

After having parameterized a pressure-corrected coarse-grained model for the binary mixture of urea and water we applied this model to parameterize benzene-urea and benzene-water interactions to study the solvation thermodynamics of benzene in urea-water at infinite dilution of benzene. As the pressure of the binary urea-water system is close to the target pressure of 1 bar we did not apply any pressure correction to the benzene-urea and benzene-water potentials; by using standard KB-IBI the benzene-urea and benzene-water KBIs were reproduced with respect to the atomistic simulations. Fig. 5.23 shows the benzene-urea and benzene-water KBIs obtained from the coarse-grained simulation and their comparison with the atomistic results. Values of all the KBIs and the derivative of the solvation free-energy of

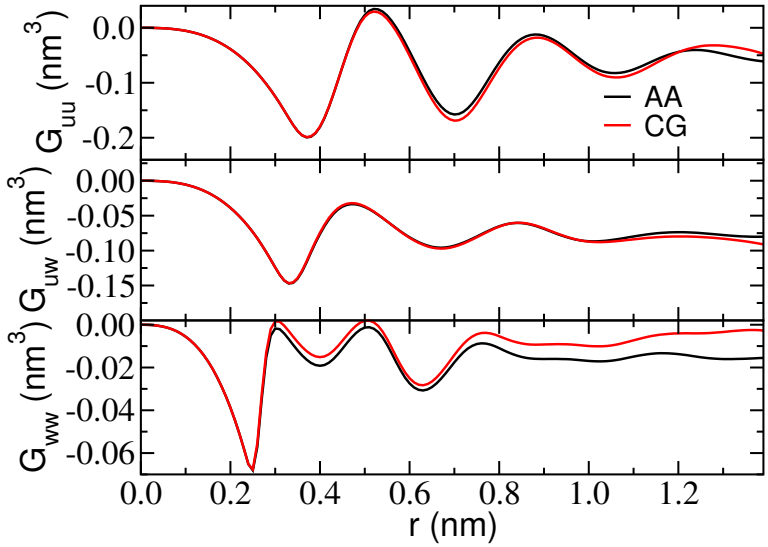


Figure 5.21: Urea-urea (uu), urea-water (uw) and water-water (ww) running KBIs ($G_{ij}(r) = 4\pi \int_0^r [g_{ij}(s) - 1]\, s^2 ds$) for binary mixture of 6 m urea and water. Shown are the results from the all-atomistic simulation (AA) and coarse-grained simulation (CG) with the pressure-corrected KB-IBI model parameterized at 6 m urea-water mixture.

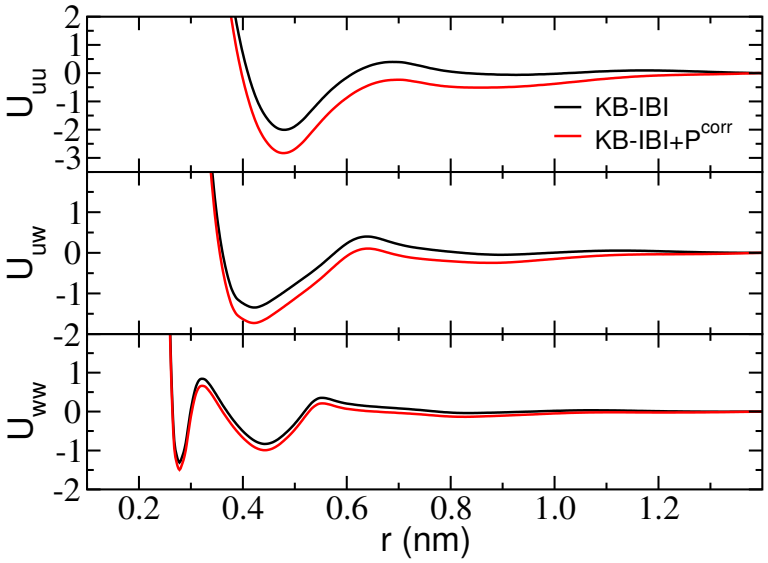


Figure 5.22: Single-site coarse-grained potentials for urea-water binary mixtures at 6 m concentration. Black lines show the KB-IBI potentials without any pressure correction to the potentials and the red lines denote the potentials where KB-IBI iterations were performed for urea-urea (uu) and urea-water (uw) potentials and simultaneously pressure correction (WJK type, see main article) was applied to the water-water potentials. All the corrections (KB-IBI and/or pressure) were preceded by 30 normal IBI iterations for all the interactions.

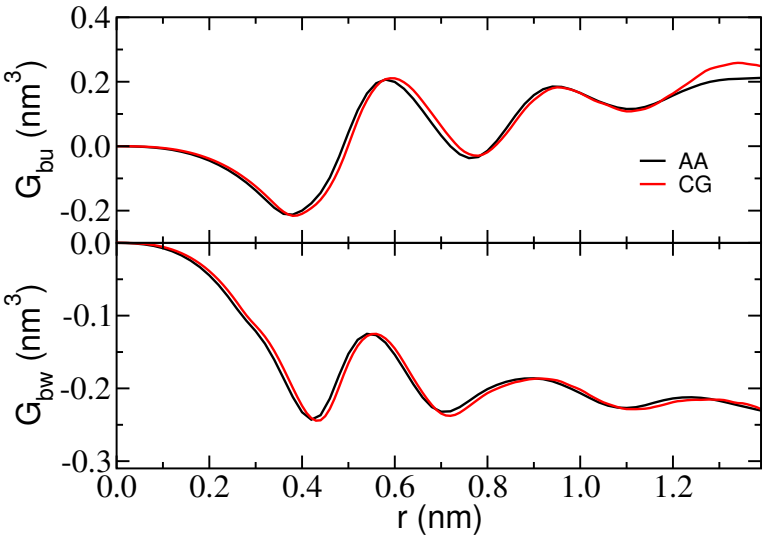


Figure 5.23: Running KB integrals $G_{ij}(r) = 4\pi \int_0^r [g_{ij}(s) - 1]\, s^2 ds$ between benzene-urea (bu) and benzene-water (bw) for a single molecule of benzene in 6 m urea-water solution. Here the potentials between benzene-urea and benzene-water are updated using KB-IBI with single benzene molecule in 6 m urea-water solution. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water with pressure correction to the water-water potential. AA: all-atomistic, CG: coarse-grained.

benzene with varying urea concentration are reported in Table 5.1 (see superscript g) and show reasonable agreement with the atomistic values irrespective of the fact that the water-water KBI is not well-reproduced. The benzene-urea and benzene-water effective pair potentials differ significantly from the corresponding potentials obtained with the binary KB-IBI solvent model of urea-water without pressure correction. The benzene-urea and benzene-water potentials for the two different solvent models are plotted in Fig. 5.24.

5.5 Conclusions

Using the recently introduced KB-IBI method,[28] we have developed single-site coarse-grained force-fields for benzene in urea-water which reproduce the preferential urea solvation and salting-in of benzene in agreement with the parent atomistic model. Using a fixed set of urea-water CG solvent potentials previously derived based on simulations of binary urea-water mixtures, solute-solvent and solute-cosolvent potentials have been newly parameterized. Solute-solute potentials for benzene at finite concentrations in urea-water solutions have further been derived in combination

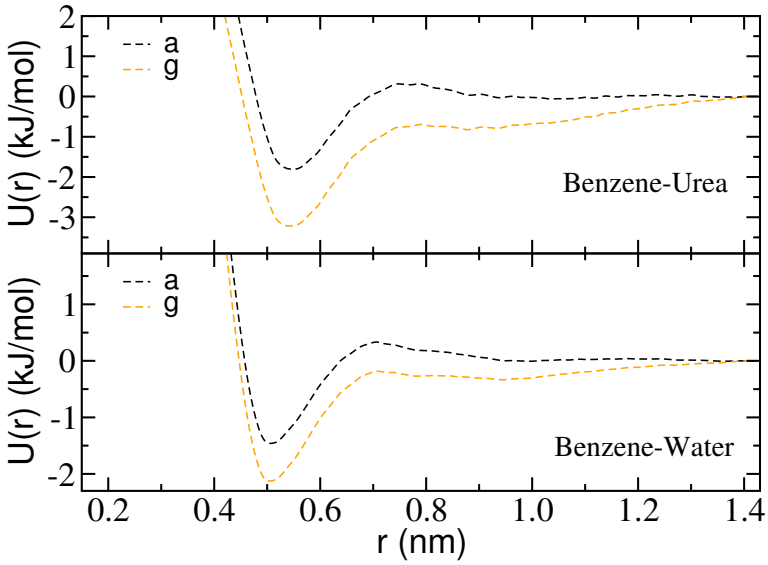


Figure 5.24: Single-site KB-IBI solute-solvent potentials used in coarse-grained molecular dynamics simulations of 6 m aqueous urea solutions with single benzene molecule. Legends denote different coarse-graining schemes as: **a**— Benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water without any pressure correction (same as in Fig. 5.2, scheme **a**). **g**— Benzene-urea and benzene-water potentials are parameterized from single benzene molecule in 6 m urea-water using KB-IBI. Urea-urea, urea-water and water-water potentials are obtained by KB-IBI method from binary mixture of 6 m urea in water with pressure correction to water-water potential (same as in Fig. 5.22, red line).

with the same CG solvent model and were used in coarse-grained MD simulations at different benzene concentrations to test the transferability of the derived models. The urea/water binary solvent model was found to be transferable to systems with finite solute concentrations as evidenced by the agreement of the solution structures and the derivatives of the benzene chemical potentials with the number of urea molecules in CG and atomistic simulations. A cluster-analysis of benzene in urea-water solutions at finite concentrations of benzene was furthermore performed based on which the free-energy of cluster growth was calculated at different urea concentrations with the detailed-atomistic and KB-IBI models. The simulations indicate that urea disfavors clustering of benzene, both, with the atomistic and KB-IBI models. The free energies of benzene clustering as predicted with the CG model were in qualitative agreement with the predictions of the detailed-atomistic model.

Inverse methods for molecular coarse graining include Inverse Monte Carlo,[47] Newton Inversion,[9] IBI,[14,15] and KB-IBI.[28] All of these methods parameterize an effective pair potential by means of iterations that aim at achieving agreement between structural target functions (usually the radial distribution functions) in the atomistic and CG systems. Because these target quantities are mostly determined by the short-range repulsive part of the effective pair potentials – with a weaker dependence on the longer-range tails – additional target quantities such as pressure and energy can be included in the parameterization. These properties are sensitive to the tails of the potentials and can, at least in principle, readily be included in the iterative procedure. Since, however, pressure and energy are not uniquely determined by the pair potential, application these thermodynamic targets in IBI or KB-IBI may cause however that properties such the isothermal compressibility or KBI are reproduced less accurately. It is shown in this work that combined application of KBI corrections and pressure corrections during the iterative optimization leads to a good compromise in reproducing the RDF, compressibility and the pressure for pure water with single-site CG models. For solution mixtures, the same can be achieved for the RDFs, KBIs and pressure. Using urea-water mixtures as model systems, it was shown that by applying pressure corrections to only the water-water pair potential during the iterative optimization of the single-site models leads to rapid convergence of the RDFs, KBIs and pressure. Moreover, since no corrections are made to any of the effective pair potentials involving the urea molecules, urea-urea, urea-water and urea-solute KBIs can be reproduced, thus keeping a consistent description of the solvation thermodynamics in systems with dissolved solutes. With

this approach, it could be shown that the salting-in behavior of benzene in urea-water can be described with single-site CG models of the solute and solvent components in good agreement with the detailed atomistic model. This shows that the KB-IBI approach to molecular coarse-graining provides models with improved properties that represent a compromise between structural and thermodynamic consistency.

Structure-based coarse-grained models are often found to be strongly state-point dependent due to inclusion of averaged multi-body contributions in the effective pair potentials. In comparison to free-energy based systematic coarse-graining methods [2] or top-down coarse-grained models such as the MARTINI model, [45] IBI-based models represent the liquid structure better by construction. The KB-IBI method improves the state-point representability in comparison to standard IBI models as it reproduces the Kirkwood-Buff integrals correctly, too. Other structure-based methods such as Newton Inversion [9] have been reported to reproduce the Kirkwood-Buff integrals of ionic liquids. [46] Although Newton Inversion or Inverse Monte Carlo methods [47] applied to complex fluids show better convergence than IBI, they involve additional computational cost and require longer simulation trajectories in order to achieve reasonable statistics needed to calculate the cross-correlations. [41] The results in this paper show that the KB-IBI method converges very rapidly when applied to solution mixtures while it provides coarse-grained binary solvent models that can readily be combined with KB-IBI models for solute-solvent and solute-solute interactions for specific solutes. The method may therefore potentially be used in future coarse-grained simulations of more complex multicomponent systems.

Bibliography

[1] Riniker, S.; Allison, J. R.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423.

[2] Brini, E.; Algaer, E.; Ganguly, P.; Li, C.; Rodriguez-Ropero, F.; van der Vegt, N. F. A. *Soft Matter* **2013**, *9*, 2108.

[3] Saunders, M. G.; Voth, G. A. *Annu. Rev. Biophys.* **2013**, *42*, 73.

[4] Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. *WIREs Comput. Mol. Sci.* **2013**, doi: 10.1002/wcms.1169

[5] Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

[6] Wang, Y.; Izvekov, S.; Yan, T.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 3564.

[7] Fritz, D.; Harmandaris, V. A.; Kremer, K.; van der Vegt, N. F. A. *Macromolecules* **2009**, *42*, 7579.

[8] Villa, A.; van der Vegt, N. F. A.; Peter, C. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2068.

[9] Lyubartsev, A.; Mirzoev, A.; Chen, L.J.; Laaksonen, A. *Faraday Discuss.* **2010**, *144*, 43.

[10] Hills Jr., R. D.; Lu, L.; Voth, G. A. *PLoS Comput. Biol.* **2010**, *6*, e1000827.

[11] Carmichael, S. P.; Shell, M. S. *J. Phys. Chem. B* **2012**, *116*, 8383.

[12] Mukherjee, B.; Delle Site, L.; Kremer, K.; Peter, C. *J. Phys. Chem. B* **2012**, *116*, 8474.

[13] Brini, E.; van der Vegt, N. F. A. *J. Chem. Phys.* **2012**, *137*, 154113.

[14] Soper, A. K. *Chem. Phys.* **1996**, *202*, 295.

[15] Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.

[16] Henderson, R. L. *Phys. Lett.* **1974**, *A49*, 197.

[17] Potestio, R. *JUnQ* **2013**, *3*, 13.

[18] Wang, H.; Junghans, C.; Kremer, K. *Eur. Phys. J. E* **2009**, *28*, 221.

[19] Fu, C.-C.; Kulkarni, P. M.; Shell, M. S.; Leal, L. G. *J. Chem. Phys.* **2012**, *137*, 164106.

[20] Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795.

[21] Schwarzinger, S.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2002**, *41*, 12681.

[22] Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5142.

[23] Stumpe, M. C.; Grubmüller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126.

[24] Trzesniak, D.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2004**, *6*, 697.

[25] Lee, M. E.; van der Vegt, N. F. A. *J. Am. Chem. Soc.* **2006**, *128*, 4948.

[26] Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15317.

[27] Canchi, D. R.; Paschek, D.; Garcia, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338.

[28] Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N. F. A. *J. Chem. Theory Comp.* **2012**, *8*, 1802.

[29] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

[30] Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.

[31] Hovorka, Ŝ.; Dohnal, V.; Carrillo-Nava, E.; Costas, M. *J. Chem. Thermodynamics* **2000**, *32*, 1683.

[32] Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, *7*, 306.

[33] Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003**, *107*, 3891.

[34] van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G.; Hochschulverlag AG an der ETH Zürich (1996).

[35] Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

[36] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

[37] Nose, S. *Mol. Phys.* **1984**, *52*, 255.

[38] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

[39] Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.

[40] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

[41] Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theo. Comput.* **2009**, *5*, 3211.

[42] Villa, A.; Peter, C.; van der Vegt, N. F. A. *J. Chem. Theory Comput.* **2010**, *6*, 2434.

[43] Ueda, M.; Katayama, A.; Kuroki, N.; Urahata, T. *Progr. Colloid & Ploymer Sci.* **1978**, *63*, 116.

[44] Ganguly, P.; van der Vegt, N. F. A. *J. Chem. Theory Comp.* **2013**, *9*, 1347.

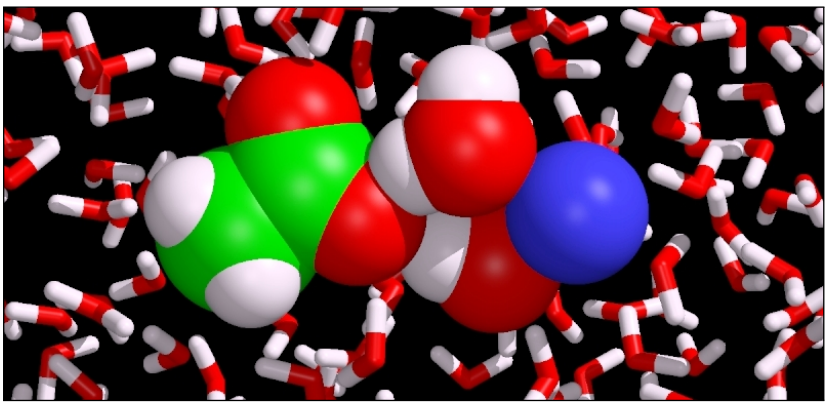
[45] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812.

[46] Wang, Y.-L.; Lyubartsev, A.; Lu, Z.-Y.; Laaksonen, A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7701.

[47] Lyubartsev, A.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730.

6 Ion Pairing in Aqueous Electrolyte Solutions with Biologically Relevant Anions

We performed molecular simulations to study ion pairing in aqueous solutions. Our results indicate that ion specific interactions of Li^+, Na^+, and K^+ with the dimethyl phosphate anion are solvent-mediated. The same mechanism applies to carboxylate ions, as has been illustrated in earlier simulations of aqueous alkali acetate solutions. Contact ion pairs play only a minor role – or no role at all – in determining the solution structure and ion specific thermodynamics of these systems. Based on the Kirkwood-Buff theory of solution we furthermore show that the well-known reversal of the Hofmeister series of salt activity coefficients, comparing chloride or bromide with dimethyl phosphate or acetate, is caused by changing from a contact pairing mechanism in the former system to a solvent-mediated interaction mechanism in the latter system.



6.1 Introduction

In the present work we shall treat the question how to relate ion specific variations in activity coefficients of aqueous electrolytes to the pairing of ions at the molecular level. The systems of interest will be the major intracellular anions (phosphates, carboxylates) relevant in biological systems. Ion specificity arises from the fact that the major effect of ions on water is short range.[1] Femtosecond pump-probe spectroscopy experiments have shown that ions have essentially no influence on the rotational dynamics of water beyond the first hydration shell.[2] First-principles molecular dynamics (MD) simulations have indicated some degree of preferential orientation of water molecules in the first hydration shell of monovalent and divalent ions, but no significant effect on water molecules located further out.[3] X-ray absorption spectroscopy (XAS) studies of cation hydration have provided a similar, local view on the effects of ions on water.[4] These observations clearly indicate that the electric fields emanating from ions in water are weak relative to the strength of water-water interactions. Therefore, forces involved in the formation of specific contact ion pairs (CIPs) can be understood only if the contribution of water molecules is explicitly considered. In this paper, thermodynamic theory[5] and molecular simulations shall be presented to illustrate the dominant role of solvent mediation in interactions between alkali cations and biologically-relevant anions. An amazingly simple and convincing concept of ion pairing has been introduced by Collins.[6,7] This concept, known as the "law of matching water affinities", considers ions, to first approximation, as charged spheres. Small ions have a high surface charge density and bind water molecules strongly (these ions are "hard" or "kosmotropic") whereas big ions bind water molecules weakly (these ions are "soft" or "chaotropic"). The discrimination between the two types comes from the relative strength of the ion-water compared to the water-water interactions. Two small ions of opposite charge experience a strong electrostatic attraction at contact distance. Although small ions bind water molecules strongly, they bind only a few of them, and the ions can come together forming direct ion pairs, expelling the hydration spheres between them. Collins assumes that two big ions can also form direct ion pairs, but through a different mechanism. Big ions bind water molecules weakly, but many of them. Release of hydration water molecules upon formation of contact pairs is favored by the relatively strong water-water interactions in these systems. The interaction between a small and an oppositely charged big ion is different: the attraction by the big

ion is not strong enough that the small ion loses its hydration shell. As a consequence a small/big ion pair is always separated by water (at low concentrations) and cannot form strong ion pairs. At large salt concentrations, small/big contact pairs can of course be formed and eventually precipitate. Collins' model has been used to successfully explain specific ion effects in colloidal and biological systems and has been confirmed by computer simulations of aqueous alkali halides by Fennell et al.[8] Hess and Van der Vegt[9] applied thermodynamic theory to simulation trajectories of aqueous alkali chlorides and could quantitatively show that ion specific changes in the salt activity coefficients are indeed caused by formation of CIPs. The major intracellular anions (phosphates, carboxylates) have been classified as "kosmotropes",[6] which according to Collins' law are expected to form contact pairs with "kosmotropic" cations. Therefore, one may assume that CIP formation with intracellular anions increases in the order $K^+ < Na^+ < Li^+$. In XAS experiments reported by Uejio et al.[10] the carbon 1s X-ray absorption transition of the carboxylate group on acetate has been studied in aqueous solutions with lithium, sodium, and potassium. The measured spectra indeed indicated monotonically stronger binding of the lighter metals. Other XAS experiments,[11] which have probed the carbonyl oxygen (rather than the carbonyl carbon) 1s X-ray absorption transition, have indicated that binding interactions with the carbonyl oxygen of acetate increase in the sequence $K^+ < Li^+ < Na^+$. The affinity of sodium over potassium for carboxyl groups on proteins has furthermore been confirmed by computer simulations.[12] These experiments and simulations are in accord with the law of matching water affinities. On a somewhat more qualitative basis, Kunz[13] has argued that the carboxyl ion is a hard ion by putting forward the reversal of the Hofmeister series of activity coefficients comparing Br^- and $CH3COO^-$ ions with this set of alkali ions. Based on computer simulations of aqueous alkali acetate solutions, Hess and Van der Vegt[9] proposed an ion pairing mechanism, which in contrast to the law of matching water affinities, does not involve the formation of CIPs. These authors showed that the number of CIPs with the carboxyl group of acetate follows the order $K^+ > Na^+ > Li^+$. This number is however small and plays no role in determining the variation of the activity coefficients among $K^+, Na^+, \text{and } Li^+$ in solutions with acetate. Instead, it could be shown that the dominant cation-carboxylate interactions are mediated through hydration water molecules at distances corresponding to the solvent-shared ion pair (SIP). The relative stability of these SIPs determines the variation of the activity coefficients and increases in the order $K^+ < Na^+ < Li^+$. Re-

cently, based on density-functional-theory (DFT) calculations of the surface charge density distributions of various ions, Dzubiella, Horinek, Netz, and coworkers[14] have shown that Br^- and the molecular anions $HCOO^-$, $CH3COO^-$, and $H2PO4^-$ are comparably "hard" or "soft". It seems, therefore, that the simple concept of "hard" and "soft" ions cannot unambiguously be applied to molecular anions and alternative mechanisms may be at play. So far, it remains unclear if the newly proposed ion pairing mechanism[9] involving water-mediated interactions is more generally valid in biological systems. To address this question, we present simulation data for alkali dimethyl phosphate (DMP) solutions and analyze ion pairing and variations of the activity coefficients. The DMP system may serve as a model for interactions of alkali cations with phosphate groups on nucleic acids.

6.2 Computational Details

All-atom simulations have been performed with the GROMACS simulation package (version 4.0).[15] We used Particle Mesh Ewald electrostatics[16] with a direct space cut off of 1.0 nm and a grid spacing of 0.12 nm. For non-bonded van der Waals interactions, a cut-off of radius 1.0 nm was used. We used the Amber force field parameters[17] of the DNA phosphate group to describe the DMP anion (using arithmetic-mean (Lorentz-Berthelot) combination rules for DMP-DMP, DMP-water, and DMP-cation interactions). The partial charges on the hydrogens were chosen to yield an overall net charge of –1. To assess the force field dependence, we also used the Gromos 43a1 force field[18] for DMP (using geometric-mean combination rules for DMP-DMP, DMP-water, and DMP-cation interactions). The force field parameters of Li^+, Na^+, and K^+ were taken from ref (9). The non-bonded van der Waals interactions between the cations (Li^+, Na^+, and K^+) and water-oxygens were obtained with geometric-mean combination rules and subsequently scaled (Table 6.1). This type of scaling has been used in previous simulations of aqueous alkali halides[9,19,20] and proved to be necessary in order to reproduce the experimental Kirkwood-Buff integrals. The scaling effectively reduces the ion-water repulsion at short-range. With unscaled Lennard-Jones parameters, the alkali cation-water interaction cannot sufficiently compensate for favorable electrostatic interactions with halide anions at short range, causing unphysical ion pairing.[9] The scaling has been applied only for the cation-water interaction, not for cation-DMP or cation-cation interactions. The SPC/E water model[21] was used in all simulations. The dielectric constant of this water model (71) is close to the experimental value (78) at room temperature,

therefore the dielectric screening of the long range electrostatics is described with reasonable accuracy. We note that the combination of the SPC/E water model with the Amber/Gromos force field is uncommon. We however point out that the hydration free energies of nonpolar and polar organic solutes are not very sensitive to the choice of the water model,[22] hence we expect that the combination of Amber/Gromos with SPC/E water provides a realistic description of the hydration of DMP. The details of the non-bonded ion parameters are given in Table 6.2. All systems contained 800 ion pairs with 44448 water molecules (1 molal concentration) in a periodic cubic box. Trajectories up to 100 ns were simulated using a MD integration time step of 1 fs. The temperature and pressure were kept constant at 298 K and 1 bar respectively using the velocity-rescale thermostat[23] and Berendsen barostat[24] respectively. All bond distances were kept constant using LINCS algorithm.[25]

Ion	σ **(nm)**	ϵ **(kJ/mol)**	**q(e)**
Li^+	0.200	0.500	1.0
Na^+	0.245	0.320	1.0
K^+	0.380	0.200	1.0

Table 6.1: Non-bonded interaction parameters of the ions

Atoms	**Scaling factor**	**Resulting** σ_{ij} **(nm)**	**Resulting** ϵ_{ij} **(kJ/mol)**
$Li^+ - O(W)$	0.40	$2.510e-01$	$2.256e-01$
$Na^+ - O(W)$	0.75	$2.778e-01$	$3.385e-01$
$K^+ - O(W)$	1.00	$3.460e-01$	$3.568e-01$

Table 6.2: Non-bonded Lennard-Jones parameters for the cation-water oxygen interaction[9] obtained after applying the scaling factor shown in the second column.

The experimental osmotic coefficients ϕ for aqueous alkali-DMP solutions reported in ref (26) have been converted to activity derivatives, analyzed further on on this work, using:

$$\phi = \left(\frac{n_S v_S + n_W v_W^g}{n_S v_W^g} \right) log\, a_W, \tag{6.1}$$

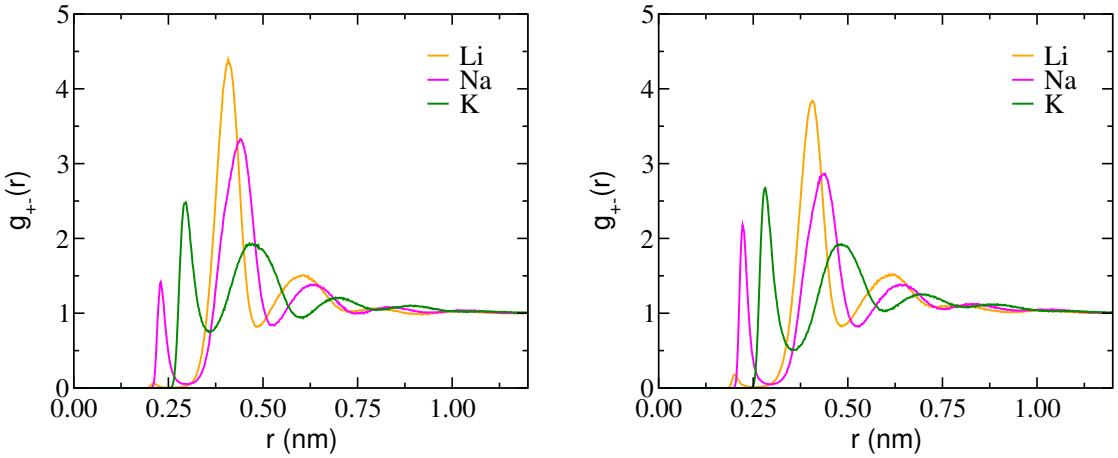
which by using a Gibbs-Duhem relation yields the required salt activity derivative

$$a_S^{'} = \left(\frac{\partial log\, a_S}{\partial log\, \rho_S} \right)_{p,T} = -\frac{n_W}{n_S} \left(\frac{\partial log\, a_W}{\partial log\, \rho_S} \right)_{p,T}. \tag{6.2}$$

In equation 6.1 and 6.2, a_W denotes the water activity, n_S and n_W denote the mole numbers salt and water, respectively. The quantities v_S and v_W^g denote the corresponding partial molar volumes, which are not known experimentally and have been taken from the simulations. For this purpose we used the salt partial molar volume at 1 M and the neat water (SPC/E) molar volume. The experimental $a_S^{'}$ values are listed in Table 6.3. Since the difference between $\mathrm{Na^+ \, and \, K^+}$ is very small (0.02) and probably within the error of computing the activity derivatives from the osmotic coefficients through equation 6.2, we shall assume that $\mathrm{Na^+ and K^+}$ behave identical. The observed experimental trend is therefore $\mathrm{Li^+ < Na^+, \, K^+}$.

6.3 Results and Discussions

We performed MD simulations of 1 molal aqueous solutions of DMP with $\mathrm{K^+, \, Na^+, \, and \, Li^+}$ cations. All simulations were performed with 800 ion pairs, while trajectories were accumulated up to 100 ns, in order to obtain good statistical accuracy of the radial distribution functions (RDFs). Figure 6.1 shows the RDFs between the cations and the (non-methylated) oxygens of DMP obtained with the Amber and m-Gromos force fields. We note that we modified the Gromos-C12 parameter for the non-methylated oxygen, as will be further explained later on. We refer to the modified Gromos model as m-Gromos. The first, second, and third peaks of the RDFs correspond to the CIP, SIP, and solvent-separated (2SIP) states. Identical trends are observed with the two force fields: the CIP state increases weight according to $\mathrm{Li^+ << Na^+ < K^+}$, while for the SIP and 2SIP states the reverse order is observed. It is moreover striking that the magnitude of the second peak for $\mathrm{Li^+ \, and \, Na^+}$ exceeds that of the first peak. Again, this is observed with both force fields. Thus, the SIP states for these systems have a lower free energy than the CIP states. It is interesting to compare the results in figure 6.1 with the corresponding data for acetate obtained previously.[9] We observe qualitatively similar changes in the SIP states by varying the cation in the systems with DMP and acetate – the SIP state with $\mathrm{Li^+}$ has the lowest free energy, followed by $\mathrm{Na^+ \, and \, next \, K^+}$. A noticeable difference between DMP and acetate involves the stability of CIPs. $\mathrm{Li^+ \, and \, Na^+}$ have a slightly stronger affinity to form CIPs with carboxylate than with phosphate. Although the implications on the thermodynamic solution properties have yet to be determined, the data in figure 6.1 clearly demonstrate the predominance of SIPs over CIPs.



(a) (b)

Figure 6.1: Alkali-dimethyl phosphate RDFs at 1 m obtained from constant pressure-temperature MD simulations (1 bar, 298 K). (a) Amber force field (b) m-Gromos force field. The RDFs are obtained by evaluating the distance between the cation and the closest of the two non-methylated oxygens of DMP; the nonspherical volume for normalization has been taken into account. The first, second, and third peaks correspond to CIP, SIP, and 2SIP states, respectively. In the SIP state, the ions are separated by a single water molecule. In the 2SIP state, two water molecules separate the two ions that are each surrounded by a complete hydration shell. Li^+ shows a very small CIP peak at 0.21 nm followed by a very high SIP peak.

Vapor pressure osmometry measurements reported by Tamaki et al. [26] have shown that the activity coefficient of aqueous alkali salts of DMP follow the order $Li^+ < Na^+ < K^+$, which is the same order as for the carboxylate-based systems. [13] A lower activity coefficient reflects stronger association between ions. Quantitative analysis of association requires taking integrals over the RDFs in figure 6.1. Below, we analyze ion association in order to address two questions. First, the analysis provides an answer to whether the simulation model reproduces the experimental observable (salt activity). Second, the analysis sheds light on the role of CIPs, SIPs, and 2SIPs in determining the order of the activity coefficients observed experimentally. To this end, we consider the derivative of the salt activity with respect to the molar salt concentration at constant pressure p and temperature T. This derivative can be related to the solution structure within the thermodynamic framework developed by Kirkwood and Buff. [5] We note that the salt activities and their concentration derivatives show the same ion specific series for aqueous solutions with chloride or acetate. [9] This also holds for DMP as could be verified by taking numerical derivatives of the

data in ref [26]. The Kirkwood-Buff (KB) relation for the concentration derivative in a two-component system reads [27]

$$a_S^{'} = \left(\frac{\partial log\, a_S}{\partial log\, \rho_S}\right)_{p,T} = \frac{1}{1 + \Delta N_{SS} - \Delta N_{WS}} \tag{6.3}$$

The salt activity is denoted with a_S, the molar salt concentration with ρ_S. The quantities ΔN_{SS} and ΔN_{WS} denote the salt-salt and water-salt excess coordination numbers defined as

$$\Delta N_{ij} = \rho_j\, 4\pi \int_0^R \left[\mathrm{g}_{ij}(r) - 1\right] r^2 dr, \tag{6.4}$$

where $g_{ij(r)}$ is the RDF for species pair ij and the limit of integration R is chosen such that $g_{ij}(r) = 1$ for $r > R$. In the KB analysis (equation 6.3 and 6.4) all ions (positive and negative) are treated as indistinguishable particles. The excess coordination number therefore includes the contributions of +/–, +/+, and –/– correlations. On the basis of our simulations, the +/– contribution will be analyzed separately below. In principle, the KB analysis can also be made considering a ternary system of water and dissociated ions. We however note that corresponding KB relations are meaningful only if the electroneutrality condition is abandoned and the three components are treated as independently fluctuating species in an open (grand canonical) system. [27] The problem which then arises is that the activity derivatives with respect to particle densities are not available from experiments. Figure 6.2 shows the quantity obtained from the simulations with the Amber and m-Gromos force fields. Bigger (i.e. more positive) values of (in the limit for large R) indicate stronger ion pairing and lower salt activity derivatives (equation 6.3). The force fields both show that DMP pairs strongest with Li^+, in agreement with the experimental data in Table 6.3 which shows the smallest value for LiDMP. The activity derivatives obtained from the simulations are also reported in Table 6.3. Quantitative agreement with the experimental data is achieved for Na^+. The best overall agreement is achieved with the m-Gromos model. The observed order, $\mathrm{Li}^+ < \mathrm{Na}^+ < \mathrm{K}^+$ for Amber and $\mathrm{Li}^+ < \mathrm{K}^+ < \mathrm{Na}^+$ for m-Gromos, is however model dependent. Experimentally, the salt activity derivatives follow the order $\mathrm{Li}^+ < \mathrm{Na}^+, \mathrm{K}^+$. While Amber seems to exaggerate the difference between Na^+ and K^+, m-Gromos closely reproduces it but with the wrong order ($\mathrm{K}^+ < \mathrm{Na}^+$). The ion specificity observed in the quantity pre-

sented in figure 6.2 results from differences in ΔN_{SS}, while ΔN_{WS} hardly depends on cation type (not shown). It is moreover clearly discernible from figure 6.2 that the important contributions to the ion specificity are due to *local* correlations at distances up to approximately 0.8 nm. Although includes all ion-ion correlations (+/+, +/–, –/–), the mutual attractions between oppositely charged ions are responsible for the ion specificity of this quantity. We therefore computed the excess coordination numbers ΔN_{+-} between oppositely charged ions, and furthermore decomposed this quantity in contributions of CIPs, SIPs and 2SIPs. In this way, molecular details of ion pairing can be related to the Hofmeister series of the salt activity.

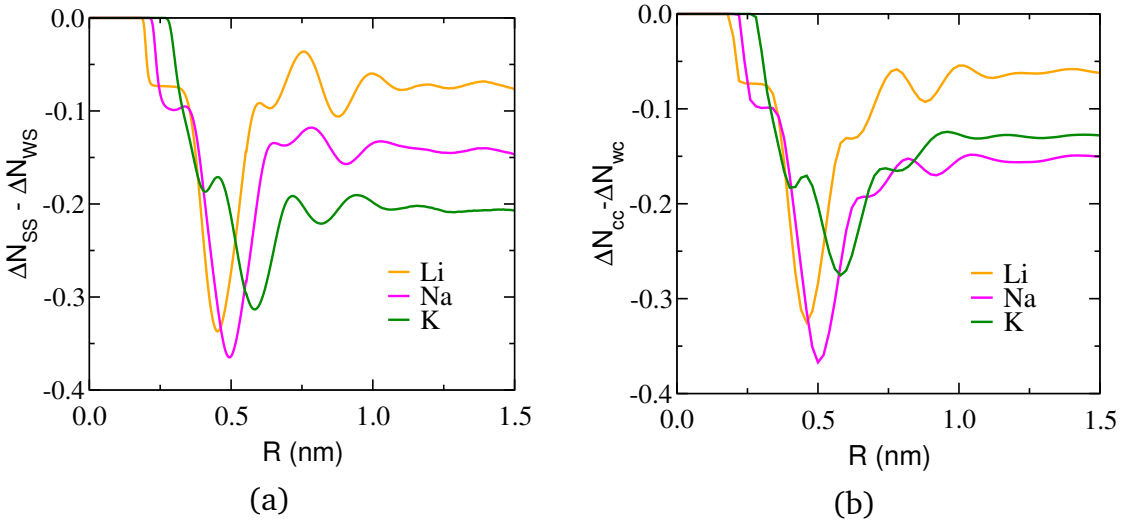


Figure 6.2: Ion specific behavior of the "salt–salt"-minus-"water–salt" excess coordination number in 1 m DMP solutions obtained from constant pressure-temperature MD simulations (1 bar, 298 K). (a) Amber. (b) m-Gromos. The excess coordination numbers are obtained from the simulations by counting ions (S) and water molecules (W) within spherical volumes of radius R around a central particle (S or W). The ion specificity of salt activity coefficients (equation 6.3) is determined by the behavior of the excess coordination numbers for large distances R (equation 6.3). In practice, however, ΔN_{SS} and ΔN_{WS} are *local* quantities with the major contributions from distances smaller than 1 nm.

In Table 6.3, we summarize the excess coordination numbers ΔN_{+-} as well as ΔN_{CIP}, ΔN_{SIP}, and ΔN_{2SIP}. The latter quantities are the spatially resolved contributions to ΔN_{+-} and are obtained by evaluating excess CIPs, SIPs, and 2SIPs in the corresponding regions. The data in Table 6.3 show that while ΔN_{+-} tends to *increase* in the order $\mathrm{K^+} < \mathrm{Na^+} < \mathrm{Li^+}$, ΔN_{CIP} tends to *decrease* (comparing $\mathrm{K+}$ and $\mathrm{Li^+}$), independent of the force field model. Hence, the variation of the salt activity derivatives with cation type ($\mathrm{K^+}, \mathrm{Na^+} > \mathrm{Li^+}$) shown in the

Ion	a_S' (exp)	a_S' (sim)	r_1	r_2	r_3	N_{CIP}	N_{SIP}	N_{2SIP}	ΔN_{CIP}	ΔN_{SIP}	ΔN_{2SIP}	ΔN_{+-}	C
AMBER													
Li^+	1.12	1.08	0.25	0.48	0.72	0.00	0.43	0.78	−0.04	0.20	0.15	0.36	0.05
Na^+	1.17	1.17	0.30	0.53	0.74	0.02	0.50	0.74	−0.05	0.21	0.11	0.35	0.07
K^+	1.19	1.26	0.36	0.60	0.80	0.09	0.58	0.79	−0.02	0.17	0.08	0.33	0.10
m-GROMOS													
Li^+	1.12	1.07	0.26	0.48	0.73	0.00	0.38	0.82	−0.04	0.15	0.17	0.38	0.10
Na^+	1.17	1.18	0.30	0.53	0.76	0.02	0.43	0.79	−0.04	0.14	0.12	0.37	0.14
K^+	1.19	1.15	0.35	0.60	0.82	0.08	0.54	0.88	−0.02	0.14	0.12	0.37	0.13

Table 6.3: **Thermodynamic and structural data of alkali DMP solutions**. Shown are: the experimental[26] activity derivatives a_S' (equation 6.1, 6.2) and the ones obtained in this work by means of equation 6.3 and 6.4 (sim.), the number of anions with in a distance r_1 (nm) around a cation (N_{CIP}), the number of anions within a distance between r_1 (nm) and r_2 (nm) (N_{SIP}) and between r_2 (nm) and r_3 (nm) (N_{2SIP}) around a cation, the corresponding excess numbers ΔN_{CIP}, ΔN_{SIP}, ΔN_{2SIP} and the overall cation-anion excess coordination number ΔN_{+-}. C denotes the long range contribution ($R > r_3$) to ΔN_{+-}. Statistical errors (obtained by block averaging) in the coordination numbers and excess coordination numbers involving the CIP, SIP, and 2SIP are all < 0.002. The statistical error in ΔN_{+-} is for all systems < 0.007.

third column of this table cannot be explained with the formation of contact ion pairs. Instead, we observe that ΔN_{SIP} and ΔN_{2SIP} follow the variation of ΔN_{+-} with cation type. We note that the excess coordination number can be written as $\Delta N_{+-} = \Delta N_{CIP} + \Delta N_{SIP} + \Delta N_{2SIP} + C$, where the constant C denotes the contribution of $R > r_3$ whose order within the set of cations is opposite to ΔN_{+-}. Thus, interactions between *hydrated* cations and *hydrated* anions determine the ion specific series of the activity coefficients in DMP solutions with lithium, sodium and potassium cations and are strongest in LiDMP solution followed by NaDMP and KDMP. The excess coordination numbers in Table 6.3 are graphically presented in figure 6.3 together with the data for alkali halides and alkali acetates obtained previously.[9] While in solution with Cl^- the trend of the excess ion coordination number (ΔN_{+-}) to grow larger for Li^+, Na^+, and K^+ can be explained by a contact ion pairing mechanism, the opposite trend for the same set of cations in acetate and DMP solutions is explained by a solvent-mediated ion pairing mechanism. Hence, the solvent-mediated interaction of Li^+ with acetate or Li^+ with DMP is stronger than that of Na^+/K^+. We finally note that in alkali halide aqueous solutions, differences between the activity coefficients and activity derivatives within this set of cations are bigger than in acetate or DMP solutions. This is likely to be caused by the more subtle variations in the strength of solvent-mediated cation-acetate/cation-DMP in-

teraction with Li^+, Na^+, and K^+ as compared to the direct contact pair interactions with halides.

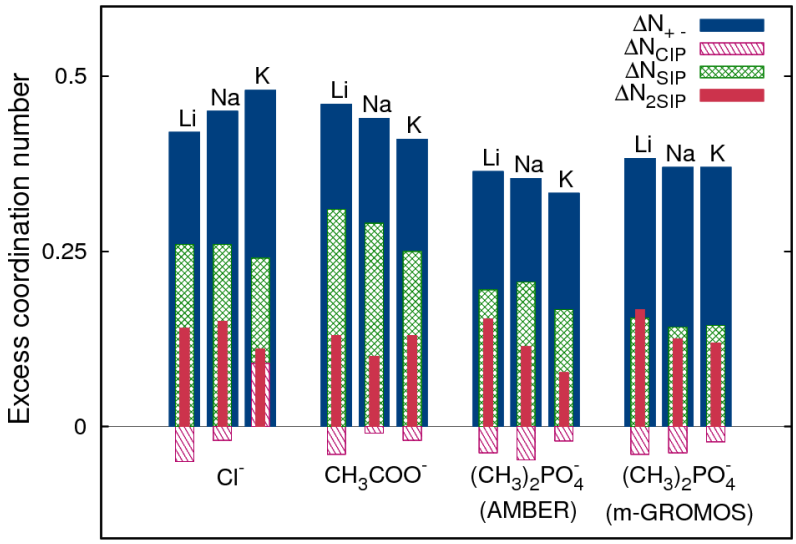


Figure 6.3: Schematic representation of excess cation-anion coordination numbers in alkali chloride[9], alkali acetate[9], and alkali DMP together with contributions of CIPs, SIPs, and 2SIPs. Cation specific variations of ΔN_{+-} are determined by CIPs, SIPs, and 2SIPs in chloride, acetate, and DMP solutions respectively.

The simulation data discussed above provide a consistent picture for two force field models (Amber and m-Gromos). Different RDFs, however, can in principle produce the same activity derivatives.[9] Therefore, it is important to ask the question if a molecular scale picture involving formation of contact ion pairs nevertheless remains possible in spite of the data presented above. We have previously demonstrated that a significant underestimation of the salt activity derivative is obtained with force field models that predict strong contact ion pairing between sodium and carboxylate ions in aqueous solution.[9] We encountered the same problem in this work when we combined the Gromos 43a1 force field model for DMP with the cation force field parameters in Table 6.1. The van der Waals C12-parameter for the nonmethylated DMP oxygen used in the Gromos 43a1 force field ($3.389 \times 10^{-6}\, kJ.mol^{-1}nm^{12}$)[18] is significantly larger than the corresponding parameter in the Amber force field ($1.590 \times 10^{-6}\, kJ.mol^{-1}nm^{12}$).[17] At the same time, the partial charge carried by this atom in the Gromos 43a1 force field (-0.6350) is smaller than in the Amber force field (-0.7761). Hence, also the charge density of this atom is smaller in comparison to the Amber model. Simulations with the Gromos 43a1 force field indicate stronger

CIP formation with K^+, as shown in Figure 6.4. This follows the expectation based on the Collins model. However, the KB analysis applied to the RDF in Figure 6.4 shows a significant underestimation of the KDMP salt activity derivative (0.99) in comparison with experiment (1.19) and the other two models in Table 6.3, i.e. formation of CIPs leads to too strong association. We therefore decided to use a smaller C12-parameter for the nonmethylated DMP oxygen in combination with the Gromos 43a1 force field. The new C12-parameter equals C12=$0.7415 \times 10^{-6}\ kJ.mol^{-1}nm^{12}$ and was taken from the first column of the interaction matrix in the Gromos force field.[18] The simulations performed with this model are referred to m-Gromos in the data that have been presented above. We are not aware of force field models that predict a contact ion pairing mechanism yet reproducing the activity derivatives.

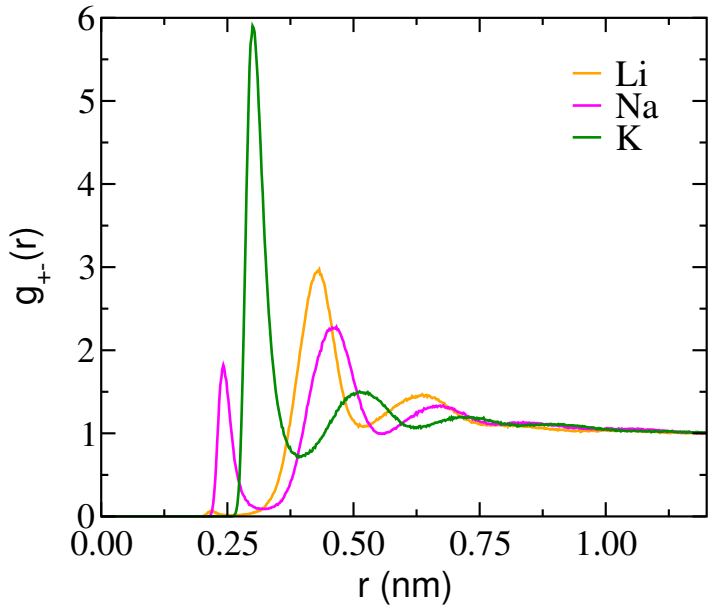


Figure 6.4: Alkali-dimethyl phosphate RDFs at 1 m obtained from constant pressure-temperature MD simulations (1 bar, 298 K) obtained with the Gromos 43a1 force field. The RDFs are obtained by evaluating the distance between the cation and the closest of the two non-methylated oxygens of DMP; the nonspherical volume for normalization has been taken into account. Li^+ shows a very small CIP peak at 0.21 nm followed by a high SIP peak.

6.4 Conclusions

Molecular simulations of ion pairing in aqueous solution are prone to force field artefacts. The molecular anions considered in this work (dimethylphosphate and acetate) have been studied with existing biomolecular force fields, which, owing to

their extensive parameterization, are assumed to provide a realistic description of the anion-water interaction. Nonetheless, validation of these models remains necessary and additional caution is needed when choosing the cation (Li^+, Na^+, K^+) force fields. Concerning the cations, it has previously been shown that models with too weak cation-water interaction predict too strong contact pair formation with anions in solution which in turn results in a significant underestimation of the salt activity derivatives.[9] In the present work, we selected force field models for cations that reproduce the salt activity derivatives of alkali chlorides and alkali acetates.[9,19] Molecular simulations were performed based on these cation force fields combined with the Amber and Gromos force fields for dimethylphosphate (DMP). The combined models were validated with respect to their ability to reproduce the experimental salt activity derivatives. Our simulations reveal that, independently of the chosen force field, the water-mediated interaction between DMP anions and lithium, sodium, or potassium cations favors formation of solvent-shared ion pairs (SIP) and solvent-separated ions pairs (2SIP) over contact ion pairs (CIP). The first peak in the lithium-DMP radial distribution function is vanishingly small, indicating that CIP states are virtually negligible in this system. The CIP/SIP balance (here defined by the relative magnitude of the peaks in the cation-anion radial distribution function) in solutions with sodium is dominated by SIP configurations, indicating that interactions between hydrated ions dominate over direct contact interactions. For potassium, the CIP peak in the RDF is slightly higher than the SIP peak. The corresponding difference in free energy (obtained from the ratio of the peak heights) is approximately $0.3\,k_BT$ in favor of the CIP interaction. A thermodynamic analysis of the structural data has furthermore been performed based on the Kirkwood-Buff theory of solutions. This analysis shows that the Hofmeister series of salt activity coefficients of aqueous alkali-DMP solutions ($Li^+ < Na^+ < K^+$) is explained by a solvent-mediated ion pair interaction mechanism (in contrast to aqueous alkali halide solutions in which contact pairs determine the thermodynamic changes[9]). A similar picture has previously been observed for alkali acetate solutions.[9] Hence, our simulation data provide support for the view that sodium and potassium interactions with intracellular anions (carboxylates and phosphates) are solvent-mediated.

6.5 Acknowledgment

The authors are grateful to Roland Netz for pointing out their theoretical calculations of ion surface charge densities in reference (14).

Bibliography

[1] collins, K. D.; Neilson, G. W.; Enderby, J. E. *Biophys. Chem.* **2007**, *128*, 95.

[2] Omta, A. W.; Kropman, M. F.; Woutersen, S.; Bakker, H. J. *Science* **2003**, *301*, 347.

[3] Krekeler, C.; Delle Site, L. *J. Phys. Cond. Mat.* **2007**, *19*, 192101.

[4] Cappa, C. D.; Smith J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J. *J. Phys. Chem. B* **2006**, *110*, 5301.

[5] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *6*, 774.

[6] Collins, K. D. *Biophys. J.* **1997**, *72*, 65.

[7] Collins, K. D. *Methods* **2004**, *34*, 300.

[8] Fennell, C. J.; Bizjak, A.; Vlachy, V.; Dill, K. A. *J. Phys. Chem. B* **2009**, *113*, 6782.

[9] Hess, B.; Van der Vegt, N. F. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13296.

[10] Uejio, J. S.; Schwarz, C. P.; Duffin, A. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6809.

[11] Ottosson, N.; Eisebitt, S.; Elberhardt, W.; Jagoda-Cwiklik, B.; Vacha, R.; Jungwirth, P.; Winter, B. *J. Phys. Chem. B* **2008**, *112*, 12567.

[12] Vrbka, L.; Vondrasek, J.; Jagoda-Cwiklik, B.; Vacha, R.; Jungwirth, P. *Proc. Natl. Acad. Sci. USA 2006*, **103**, 15440.

[13] Kunz, W. *Curr. Opin. Coll. Intf. Sci.* **2010**, *15*, 34.

[14] Dzubiella, J.; Fyta, M.; Horinek, D.; Kalcher, I.; Netz, R. R.; Schwiertz, N. In *Specific Ion Effects*; Kunz, W. Ed.; World Scientific: Singapore. 2010; Chapter 9.

[15] Hess, B.; Kutzner, C.; van der Spoel, D., Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

[16] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

[17] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

[18] Van Gunsteren, W. F.; Billeter, S. R., Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G., *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; vdf Hochschulverslag: ETH ZÃ¼rich, Switzerland, 1996.

[19] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *119*, 11342.

[20] Klasczyk, B.; Knecht, V. *J. Chem. Phys.* **2010**, *132*, 024109.

[21] Berendsen, H. J. C.; Grigera, J. R., Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

[22] Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616.

[23] Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

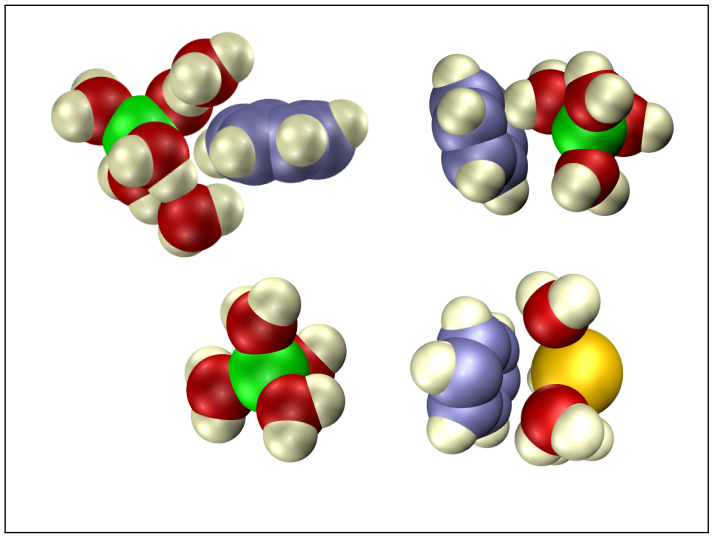
[24] Berendsen, H. J. C.; Postma, J. P. M. Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

[25] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comp. Chem.* **1997**, *18*, 1463.

[26] Tamaki, K.; Suga, K.; Tanihara, E. *Bull. Chem. Soc. Jpn.* **1987**, *60*, 1225.

[27] Ben-Naim, A. *Molecular Theory of Solutions*; Oxford Univ. Press, New York (2006).

7 Salting-out of Benzene with Hofmeister Cations: an Insight to the Hydrophobic Ring-Cation Interactions



We study the ion-specific salting-out process of benzene in aqueous alkali chloride solutions using Kirkwood-Buff theory of solutions and molecular dynamics simulations with different force-fields for the ions and benzene. The differences between the Setchenow salting-out coefficients of the different salts are found to be determined by the direct interactions between the benzene molecules and the cations whereas the propensity of the salt ions to form neutral ion pairs in water do not play any significant role. Our simulations do not show any cation-specific effect on bulk water structure and benzene-water correlations. The analysis of ion and water distributions around benzene molecules further revealed direct interactions of the bigger chaotropic cations (K^+, Rb^+, Cs^+) with benzene along the vertical axis to the benzene-ring and weak water-mediated interactions with benzene horizontally. Kosmotropic Li^+ showed relatively stronger water-mediated interactions vertically and horizontally to the benzene-ring, thus providing an explanation for the deviating position of lithium close to rubidium in the experimental salting-out series. The interactions between lithium ion and benzene are found to be insensitive to the small partial charges on the atoms of the benzene molecules and it proposes a qualitative explanation for the smaller salting-out co-

efficient of lithium chloride than sodium or potassium chloride for salting out small pure hydrophobic molecules. The present work illustrates that Kirkwood-Buff fluctuation theory provides a general framework, based on which direct interactions and indirect effects related to solvent non-idealities can be studied in computer simulations of salting-in and salting-out phenomena relevant to questions of biomolecular stability.

7.1 Introduction

Ion-specific interactions with proteins were studied back in 19th century by Hofmeister.[1] Since then several experimental studies have been carried out to rank the ions/salts depending on their ability to salt-in or salt-out solutes in aqueous solutions. Ion-specific Hofmeister effects lie at the basis of several phenomena observed in bulk solution and at interfaces,[2–7] and have intensively been studied in the last decade.[8,9] Although the relative effects of ions on properties such as viscosity, density, heat capacity, activity coefficient and osmotic pressure of a solution are well-characterized in experiments, the underlying mechanisms for the ion-specificity at molecular level are not strongly established. There has been debate over the years as to whether salting-in or salting-out effects of ions are related to effects these ions have on bulk water properties. Recent consensus however invokes the idea that direct binding interactions of ions with water-protein interfaces or with chemical groups on macromolecules in water determine the salting-in and salting-out effects.[6,7,10–13]

Recently there have been several theoretical studies on hydrophobic solutes in aqueous salt solutions. Zangi *et al.*[14] studied the interactions between two hydrophobic plates in water with model ions with increasing electronic charge-density and found that salting-out, attributed to ions with higher charge-density, is purely entropic and on the contrary, salting-in, caused by ions with lower charge-density, can be either entropic or enthalpic. The correlation between the increase in hydrophobic interaction and the increase in ion-charge-density was also found in early studies.[15,16] But in many systems of hydrophobes with electrolytes, Li^+ salts do not follow the expected order according to the charge-density.[3,17,18] For example, the Setchenow constant, which is defined as the slope of the relative solubility curve of a solute with the variation in the salt concentration,[19] is similar for Li^+ and Rb^+ when considering benzene in aqueous alkali chlorides.[3,17] In theoretical simulations of small hydrophobic solutes, methane and neopentane, in aqueous alkali halide solutions, Thomas and Elcock found a correlation between the hydrophobic interactions and experimental solubility data which further correlated with the change in the hydrogen-bonded structure of water.[20] In a later work Graziano used scaled particle theory and calculated the reversible work for cavity formation in aqueous solutions of benzene with alkali chlorides and found that the solubility of benzene is determined by the volume packing density and the effective hard-sphere diam-

eter.[21] Athawale *et al.* performed simulations of small hydrophobic solutes and larger hydrophobic chain molecules in aqueous NaCl solutions and found that enthalpy dominates the salting-out behavior for small hydrophobes whereas for larger hydrophobic chains entropy dominates the hydrophobic folding of the chains.[22] In an experimental study of aqueous mixtures of ionic liquids and sodium salts Tomé *et al.* showed that salting-out is entropic but salting-in occurs because of the direct binding of the ions with the hydrophobic groups of the ionic liquids.[23]

In this paper we study the ion specific salting-out of benzene from water by alkali chloride salts. We use molecular dynamics (MD) simulations combined with fluctuation theory of solutions (also referred to as Kirkwood-Buff theory).[24] The combination of simulations and fluctuation theory offers the advantage of providing an easy and exact route to relate the thermodynamic Setchenow coefficients to the structural correlations at pair level. Using this approach, we will show that differences between the Setchenow coefficients of different salts cannot be determined by the differences in ion pairing propensities at the atomic scale rather the differences are caused by the interactions between benzene and the cations. We will also show that lithium shows a comparatively stronger direct interaction with benzene mediated by water bridges, causing lithium chloride to be less salting-out agent than sodium or potassium chloride. Also we will find that for all the cases benzene-water and water-water pair correlations (water structure) play no role in ion-specificity of the salting-out phenomena.

7.2 Thermodynamic Theory

We use the statistical mechanics theory of Kirkwood and Buff,[24] which relates integrals over molecular distribution functions to particle number fluctuations in open systems. These fluctuations determine thermodynamic response functions and derivatives of chemical potentials with respect to concentrations. An analysis of these fluctuations in closed systems allows to obtain the Setchenow salting-out coefficient, which is proportional to the derivative of the solute (benzene) solvation free energy (ΔG_b) with respect to the salt concentration (mole fraction) x_{s},[25]

$$\lim_{\rho_{\mathrm{b}}\to 0}\left(\frac{\partial \Delta G_{\mathrm{b}}}{\partial x_{\mathrm{s}}}\right)_{p,T} = \frac{RT\left(\rho_{\mathrm{w}}+\rho_{\mathrm{s}}\right)^2}{\eta}\left(\mathrm{G_{bw}}-\mathrm{G_{bs}}\right), \tag{7.1}$$

or with respect to the salt molarity c_{s},

$$\lim_{\rho_{\mathrm{b}}\to 0}\left(\frac{\partial \Delta G_{\mathrm{b}}}{\partial c_{\mathrm{s}}}\right)_{p,T}=\frac{RT\left(\rho_{\mathrm{w}}+\rho_{\mathrm{s}}\right)^{2}}{\eta}\left(\mathrm{G_{bw}}-\mathrm{G_{bs}}\right)\left(\frac{\partial x_{\mathrm{s}}}{\partial c_{\mathrm{s}}}\right), \tag{7.2}$$

where subscript b stands for benzene, s for salt and w for water. R is the gas constant, p denotes pressure and T the temperature. ρ_b, ρ_w and ρ_s are the molar concentrations of benzene, water and salt, respectively, and $\mathrm{G_{bw}}$ and $\mathrm{G_{bs}}$ are the benzene-water and benzene-salt Kirkwood-Buff integrals which will be defined below. Eq. 7.1 is exact in the limit of zero solute concentration. The positive constant η is given by

$$\eta=\rho_{\mathrm{w}}+\rho_{\mathrm{s}}+\rho_{\mathrm{w}}\rho_{\mathrm{s}}\left(\mathrm{G_{ww}}+\mathrm{G_{ss}}-2\mathrm{G_{sw}}\right). \tag{7.3}$$

The quantities $\mathrm{G}_{\alpha\beta}$ are the Kirkwood-Buff integrals (KBIs) defined as

$$\mathrm{G}_{\alpha\beta}=4\pi\int_{0}^{\infty}\left[\mathrm{g}_{\alpha\beta}(r)-1\right]r^{2}dr, \tag{7.4}$$

where $\mathrm{g}_{\alpha\beta}(r)$ is the radial distribution function (RDF) between solution components α and β. For systems away from instability points, contributions to the integral in Eq. 7.4 are usually local and determined by fluctuations within distances $r<1.0-1.5$ nm.[26] Physically, KBIs provide a measure for the preferential solvation or relative affinity between the solution components. Hence, if $(\mathrm{G_{bw}}-\mathrm{G_{bs}})>0$ the benzene-water affinity exceeds the benzene-salt affinity, which, according to Eq. 7.2, must lead to salting-out; *i.e.* $(\partial\Delta G_{\mathrm{b}}/\partial c_{\mathrm{s}})_{p,T}>0$.

In the present study, we compute Eq. 7.2 at 1 molal salt concentration for aqueous alkali chloride solutions with $\mathrm{Li^{+}}$, $\mathrm{Na^{+}}$, $\mathrm{K^{+}}$, $\mathrm{Rb^{+}}$ and $\mathrm{Cs^{+}}$ ions. The experimental $(\partial\Delta G_{\mathrm{b}}/\partial c_{\mathrm{s}})_{p,T}$ values can be calculated by using the experimental Setchenow constants k_{s}[17] as[21]

$$\left(\frac{\partial \Delta G_{\mathrm{b}}}{\partial c_{\mathrm{s}}}\right)_{p,T}^{exp}=2.3\,RT\,k_{\mathrm{s}}\,. \tag{7.5}$$

We note that in the analysis of the simulation data, the cation and anion are treated indistinguishably, *i.e.* KBIs (Eq. 7.4) involving the salt are obtained from pair corre-

lation functions $g_{\alpha s}(r)$ in which cations and anions are treated as indistinguishable particles.

7.3 Computational Details

Molecular simulations were performed using the GROMACS molecular dynamics package, version 4.0.7.[27] Gromos43a1[28] and Gromos53a6[29] force field parameters were used for benzene. For water the SPC/E model[30] was used. For the ions we used four sets of parameters, namely 1. Smith (Li^+, K^+, Rb^+, Cs^+ and Cl^-),[31,32] 2. Hess (Li^+, K^+ and Cl^-),[33] 3. Netz (Na^+, K^+, Cs^+ and Cl^-),[34] 4. Dang (Li^+ and Cl^-).[35] Smith and Hess models are the Kirkwood-Buff derived force fields with an appropriate scaling for cation-water interactions. Netz model uses ion-water interaction parameters from single-ion solvation data and appropriate scaling factors for cation-anion interactions are used to reproduce experimental Kirkwood-Buff integrals (for Cl^- no scaling factor is used). Smith, Hess and Netz ion models combined with the SPC/E water model reproduce the experimental Kirkwood-Buff integrals and activity-coefficients of the corresponding electrolyte solutions. All the non-bonded interaction parameters for benzene, ions and water are listed in table 7.1. NpT simulations were performed at a temperature of 298 K and a pressure of 1 bar using a velocity-rescale thermostat[36] with a relaxation time of 0.1 ps and the Berendsen barostat[37] with a coupling time of 1 ps. All bonds of the benzene molecules were constrained using the SHAKE algorithm.[38] The particle mesh Ewald (PME)[39] method was used for calculating the electrostatic interactions with a grid-spacing of 0.12 nm. The cut-off radius for all non-bonded interactions was 1 nm. The Newtonian equations of motion were integrated using a leap-frog integrator with a 2 fs time-step. Trajectories were accumulated up to 150 ns. Periodic cubic boxes with average linear box dimensions of ~ 11 nm containing $\sim$ 44444 water molecules, 800 ion-pairs (1 molal concentration) and 25 benzene molecules were simulated. Large simulation boxes are required to achieve a low benzene concentration with a sufficient number of benzene molecules in order to guarantee accurate statistics of the benzene-water and benzene-salt correlations. Kirkwood-Buff integrals were calculated by taking the average of the running KBIs ($G_{ij}(r) = 4\pi \int_0^r [g_{ij}(s) - 1]\, s^2 ds$, r being the radial distance) between 0.9 and 1.2 nm.

Benzene						
	σ_{C}	ϵ_{C}	σ_{H}	ϵ_{H}	q_{C}	q_{H}
43a1[28]	0.3361	0.4059	0.2373	0.1184	−0.10	0.10
53a6[29]	0.3581	0.2774	0.2373	0.1184	−0.14	0.14

Ion													
	σ_{Li}	ϵ_{Li}	σ_{Na}	ϵ_{Na}	σ_{K}	ϵ_{K}	σ_{Rb}	ϵ_{Rb}	σ_{Cs}	ϵ_{Cs}	σ_{Cl}	ϵ_{Cl}	Comb.
Smith[32]	0.1820	0.7000	0.2450	0.3200	0.3340	0.1300	0.3620	0.1500	0.4130	0.0650	0.4400	0.4700	1
Hess[33]	0.2000	0.5000	–	–	0.3800	0.2000	–	–	–	–	0.4400	0.4700	1
Netz[34]	–	–	0.2583	0.4186	0.2690	2.4400	–	–	0.3331	1.5400	0.4400	0.4186	2
Dang[35]	0.1506	0.7000	–	–	–	–	–	–	–	–	0.4400	0.4186	2

Cation-Water										
	$\sigma_{\mathrm{Li-O}}$	$\epsilon_{\mathrm{Li-O}}$	$\sigma_{\mathrm{Na-O}}$	$\epsilon_{\mathrm{Na-O}}$	$\sigma_{\mathrm{K-O}}$	$\epsilon_{\mathrm{K-O}}$	$\sigma_{\mathrm{Rb-O}}$	$\epsilon_{\mathrm{Rb-O}}$	$\sigma_{\mathrm{Cs-O}}$	$\epsilon_{\mathrm{Cs-O}}$
Smith[32]	0.2400	0.2700	0.2785	0.3420	0.3252	0.2327	0.3385	0.2655	0.3616	0.1954
Hess[33]	0.2516	0.2281	–	–	0.3469	0.3607	–	–	–	–
Netz[34]	–	–	0.2876	0.5216	0.2930	1.2600	–	–	0.3250	1.0000

Water						
	σ_{O}	ϵ_{O}	σ_{HW}	ϵ_{HW}	q_{O}	q_{HW}
spc/e[30]	0.3166	0.6506	0.0	0.0	−0.8476	0.4238

Table 7.1: Non-bonded interaction parameters for ions, benzene and water. Comb.: combination rule. Combination rule 1: $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$, $\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$; combination rule 2 (Lorentz-Berthelot): $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$, $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$. Except for cation-oxygen(water) interactions, all the other interactions are governed by the combination rules. For Smith and Hess force-fields ϵ values of the cation-oxygen(water) interactions are scaled as– Li^+: 40%, Na^+: 75%, K^+: 80% for Smith force-fields and 100% for Hess force-fields, Rb^+: 85%, Cs^+: 95%. The final cation-oxygen(water) ϵ values are reported. For Netz force-fields separate cation-oxygen(water) interaction parameters are used and reported accordingly. All the σ and ϵ values are in nm and kJ/mol respectively. q-s are the electronic charges on the atoms; all the cations have charges $+1.0$ and Cl^- has a charge -1.0. O: oxygen(water), HW: hydrogen(water).

The benzene-cation RDFs, obtained using gromos53a6 parameters for benzene and Smith force-fields for ions, are plotted in Fig. 7.1. Clearly, the benzene-Li^+ correlations are stronger than those involving the other cations. The large cations (K^+, Rb^+, Cs^+) show a peak at small distances ($< 0.4\,nm$) while this peak is missing for the small cations (Li^+, Na^+). Li^+ shows a broad peak at 0.55 nm preceded by a shoulder at 0.45 nm. To obtain a more detailed picture, we further studied the distributions of the cations around the benzene ring considering three directions: 1) the in-plane direction along the C-H bonds (X-axis), 2) the in-plane direction orthogonal to the C-C bonds (Y-axis) and 3) the direction perpendicular to the ring (Z-axis), the center of mass of benzene being the origin in all cases. The distributions along Z and X are shown in Fig. 7.2. From the distributions of the cations along Z (from top), we observe that the large cations form contact pairs with benzene from the top which explains the first peaks of the benzene-cation RDFs. The small Li^+ and Na^+ ions do not form contact pairs with benzene from the top, instead they form a water-mediated pair with benzene from the top, which explains the shoulder in the benzene-cation RDFs (for Li^+ at $\sim$ 0.45 nm and for Na^+ at $\sim$ 0.5 nm). The broad peak of the benzene-Li^+ RDF is further explained by the distribution of Li^+ ions that approach the benzene ring from the sides (along X and Y). Along X (also along Y, data not shown) we find that the Li^+ distribution exhibits a maximum at $\sim$ 0.5 nm which is absent for the large cations. A simulation snapshot of Li^+ at this distance (shown in Fig. 7.3a) shows that the benzene molecule penetrates the first solvation shell of Li^+ equatorially without modifying the hydration shell structure of Li^+ and finds an energetically favorable arrangement. The coordination number of the water-oxygens around Li^+ remains 4 in these instances which is same as in bulk water where no benzene molecule approaches Li^+. From the analysis of the distributions of the cations around benzene it appears that a water-mediated interaction from the top of the benzene ring and a direct approach towards benzene from the side contribute to a relatively larger affinity of Li^+ for benzene in comparison to the other cations. We have tested three different force fields for Li^+, one using the parameters developed by Gee *et al.*[32] (Smith force-fields, data shown) and the other ones using the parameters developed by Hess and van der Vegt[33] and Dang[35] (Hess and Dang force-fields respectively, data not shown). All the force fields showed very similar results in terms of the coordinations of the lithium ions

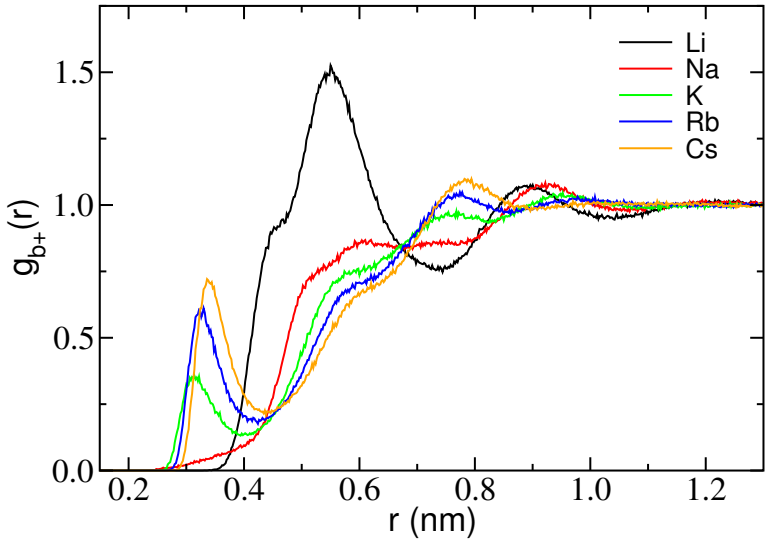


Figure 7.1: Radial distribution functions between benzene and cations in 1 m aqueous solutions of alkali chlorides. Force-fields used– benzene: gromos53a6; ions: Smith.

around the benzene molecules. A few more illustrative snapshots showing different mechanisms of the cations (Li^+, Cs^+) interacting with benzene, and the first hydration shell of Li^+ in bulk water, are presented in Fig. 7.3. The RDFs between benzene and Cl^- (see Fig. 7.4) showed very similar results for all salts except LiCl which showed a slightly higher affinity of Cl^- around benzene.

In order to link the local correlations between solution components to salting-out thermodynamics we calculated the KBIs appearing in Eq. 7.2 and Eq. 7.3. Fig. 7.5 shows the quantity $[G_{bw}(r) - G_{bs}(r)]$ (obtained using gromos53a6 for benzene and Smith parameters for ions), whose limiting value for large r equals the term $(G_{bw} - G_{bs})$ in Eq. 7.2. For all salts we observe $(G_{bw} - G_{bs}) > 0$, *i.e.* the simulations predict preferential hydration and salting-out of benzene by alkali chloride salts in agreement with experiments. The experimental Setchenow salting-out coefficient follows the order $Na^+ > K^+ > Rb^+$, $Li^+ > Cs^+$.[3] The running KBIs that characterize the affinities between solvent components (salt-salt, salt-water and water-water) are plotted in Fig. 7.8. These KBIs determine the constant η in Eq. 7.2. From Eq. 7.2 it is clear that the quantity $(\partial \Delta G_b / \partial c_s)_{p,T}$ is approximately proportional to $(G_{bw} - G_{bs})$ and inversely proportional to η as the quantity $(\rho_w + \rho_s)$ is dominated by ρ_w (at comparatively low salt-concentrations) and does not vary significantly with different salts studied. So the salting-out order is determined only by $(G_{bw} - G_{bs})$ and η. We first start with analyzing the term η which includes the indirect correlations of the solutions (with respect to the solute), namely salt-salt,

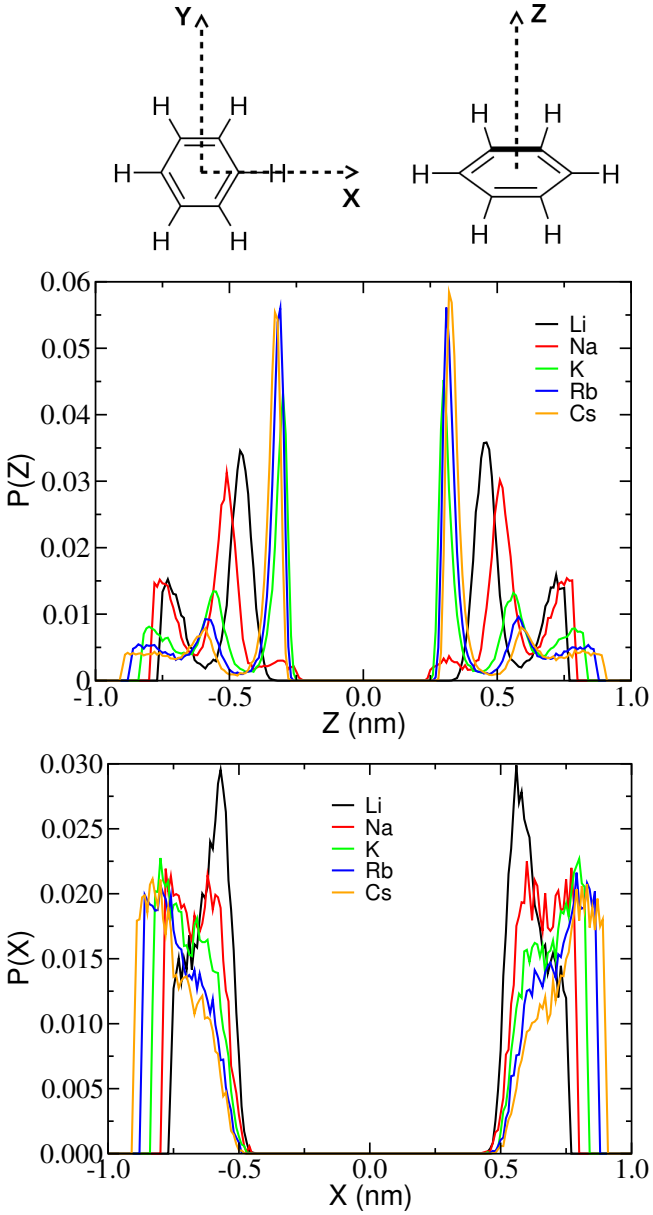


Figure 7.2: Distribution of the cations around benzene molecules from the top of the ring (upper panel) and along $C-H$ bonds (lower panel). The cut-off for the distributions was chosen to be the length at which the first minimum after the second prominent peak of benzene-cation RDFs occurs (in nm: Li^+ 0.76, Na^+ 0.79, K^+ 0.83, Rb^+ 0.87, Cs^+ 0.90). Force-fields used– benzene: gromos53a6; ions: Smith.

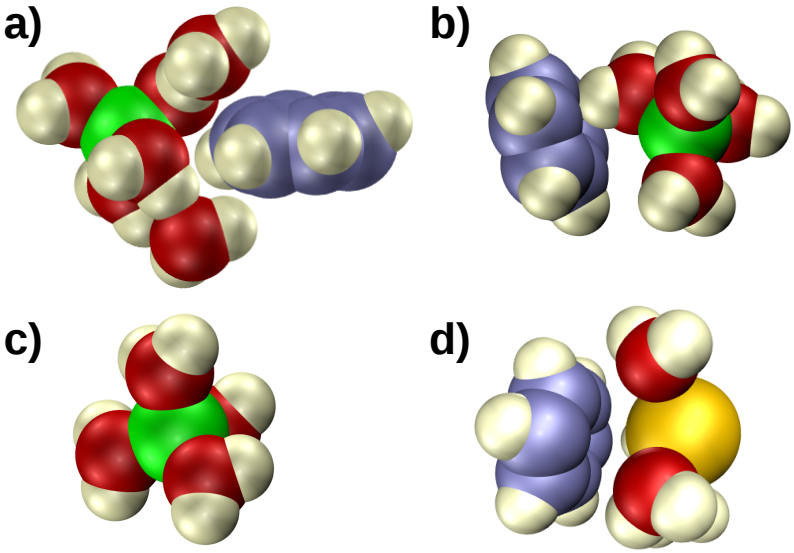


Figure 7.3: Snapshots illustrating different interaction mechanisms of cations with benzene and water. a) Li^+ approaching benzene from the side. Benzene can penetrate the hydration shell of Li^+ without perturbing it significantly. b) water-mediated interaction of Li^+ with benzene. Benzene π-electrons accept a hydrogen bond from a water molecule in the hydration shell of Li^+. c) Hydrated Li^+ ion in bulk water. d) Direct interaction of Cs^+ with benzene from the top. Color-codes: $\mathrm{green - Li^+,\ orange - Cs^+,\ grey - H,\ red - O,\ iceblue - C}$.

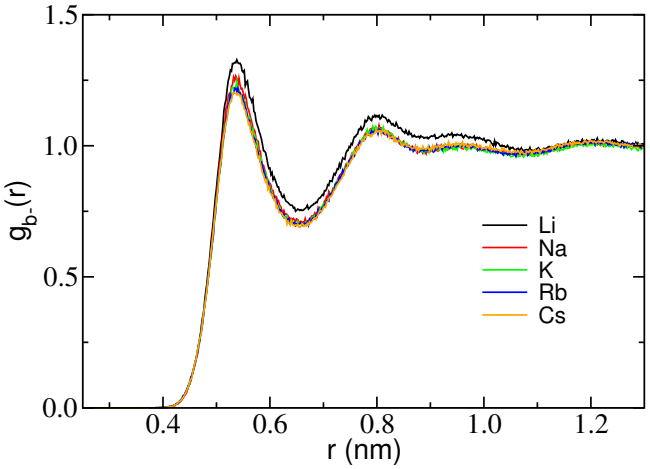


Figure 7.4: Radial distribution functions between benzene and chloride ion (Cl^-) in 1 m aqueous solutions of alkali chlorides with very dilute concentration of benzene. Force-fields used– benzene: gromos53a6; ions: Smith.

salt-water and water-water correlations which are the properties of the binary mixtures of salts and water without the solute particles inserted. For different force-fields used for benzene and the ions η values are listed in table 7.2. Though we calculated the η values from ternary mixtures of benzene, salt and water, these η-s still reflect the properties of binary salt-water mixtures approximately as benzene concentration is too low. We find that different Kirkwood-Buff derived ion-parameters (Smith[32], Hess[33] and Netz[34]) yield very similar salt-salt (Hess force-fields for KCl show a relatively higher salt-salt aggregation), salt-water and water-water KBIs for individual cations which are also reported to be very close to the experimental KBIs for binary salt-water mixtures[32–34]. So one can argue that the η values calculated in our studies are in reasonable agreement with experiments. If we compare the η values for different cations, we do not find significant variation in η with different cations which would reflect in the significant variation in $\left(\partial \Delta G_{\mathrm{b}}/\partial c_{\mathrm{s}}\right)_{p,T}$. Experimentally $\left(\partial \Delta G_{\mathrm{b}}/\partial c_{\mathrm{s}}\right)_{p,T}$ values (see table 7.2) vary from $1.11\ kJ/mol$ (for Na^{+}) to $0.50\ kJ/mol$ (for Cs^{+}) which is a variation by a factor of 2.22 where as the η values vary from $\approx 33\ nm^{-3}$ (for Na^{+}) to $\approx 34\ nm^{-3}$ (for Cs^{+}). Clearly the variation in $\left(\partial \Delta G_{\mathrm{b}}/\partial c_{\mathrm{s}}\right)_{p,T}$ for different cations does not result from the variation in η. So one can argue that the variation in the salting-out coefficient for the salts of the different cations is due to the variation in the term $(\mathrm{G_{bw}} - \mathrm{G_{bs}})$ which deals with the direct correlations involving the solute. But when Smith force-fields for ions are combined with gromos53a6 force-fields for benzene it is found that (Fig. 7.5) the values of $(\mathrm{G_{bw}} - \mathrm{G_{bs}})$ are almost identical for all the cations except Li^{+} if we take the average of the running KBIs between 0.9 to 1.2 nm (beyond this distance KBIs suffer from convergence issues and finite-size effects[26]). In fact, Na^{+} shows a slightly lower value of $(\mathrm{G_{bw}} - \mathrm{G_{bs}})$ than the other cations (except Li^{+}) which is opposite to the salting-out order found. This, combined with η, results in very similar salting-out coefficients for $\mathrm{Na}^{+},\ \mathrm{K}^{+},\ \mathrm{Rb}^{+}$ and Cs^{+}. $\left(\partial \Delta G_{\mathrm{b}}/\partial c_{\mathrm{s}}\right)_{p,T}$ values are reported in table 7.2 which do not show satisfactory agreement with the experimental results except for Rb^{+}. To study the effect of the different benzene models we computed the KBIs using Smith force-field for ions combined with gromos43a1 model for benzene (gromos53a6 benzene model shows significantly lower excess chemical potential of solvation in spc water[41] than gromos43a1 benzene model and experimental results[42]) but no significant difference in the KBIs was found. Netz parameters for ions ($\mathrm{Na}^{+},\ \mathrm{K}^{+}$ and Cs^{+}) in combination with gromos43a1 benzene show correct salting-out order between $\mathrm{Na}^{+},\ \mathrm{K}^{+}$ and Cs^{+} ions but quantitatively do not match

with experiments although the derivative of free-energy of solvation of benzene for Cs^+ is predicted significantly closer to the experiments than that predicted by Smith parameters due to a stronger benzene-salt interactions. All the corresponding KBIs are listed in table 7.2. A stronger benzene-Cs^+ direct interaction for Netz parameters can be observed from the benzene-Cs^+ RDFs plotted in Fig. 7.9. The quantity, $G_{bw} - G_{bs}$, obtained using Netz parameters for ions and gromos43a1 benzene model, is also plotted in Fig. 7.10 for different cations. From table 7.2 and Fig. 7.7 we find that the benzene-water KBIs do not change significantly with different cations or different force-fields. Our results show strong indications that the variation in the salting-out coefficients for different cations is due to the change in the benzene-salt correlations which the present models do not reproduce accurately.

Cation	Force-field	G_{bw}	G_{bs}	G_{ss}	G_{sw}	G_{ww}	η	$\left(\frac{\partial \Delta G_b}{\partial c_s}\right)^{sim}_{p,T}$	$\left(\frac{\partial \Delta G_b}{\partial c_s}\right)^{exp}_{p,T}$
Li^+	Smith+53a6	−0.124	−0.392	−0.069	−0.014	−0.029	31.62	0.42	0.80
	Smith+43a1	−0.115	−0.355	−0.072	−0.015	−0.029	31.61	0.37	
	Hess+53a6	−0.126	−0.346	−0.078	−0.016	−0.029	31.50	0.34	
	Dang+43a1	−0.118	−0.435	−0.087	−0.017	−0.029	31.31	0.50	
Na^+	Smith+53a6	−0.131	−0.648	−0.008	−0.010	−0.029	32.81	0.78	1.11
	Smith+43a1	−0.118	−0.629	−0.010	−0.010	−0.029	32.80	0.77	
	Netz+43a1	−0.113	−0.650	0.005	−0.017	−0.029	33.10	0.79	
K^+	Smith+53a6	−0.128	−0.670	0.040	−0.021	−0.029	33.76	0.79	0.95
	Smith+43a1	−0.117	−0.662	0.037	−0.021	−0.029	33.73	0.80	
	Hess+53a6	−0.119	−0.662	0.071	−0.034	−0.028	34.40	0.76	
	Netz+43a1	−0.114	−0.635	0.042	−0.020	−0.029	33.83	0.76	
Rb^+	Smith+53a6	−0.115	−0.656	0.061	−0.027	−0.029	34.16	0.78	0.80
	Smith+43a1	−0.105	−0.676	0.052	−0.027	−0.028	34.03	0.83	
Cs^+	Smith+53a6	−0.122	−0.653	0.050	−0.034	−0.028	33.98	0.77	0.50
	Smith+43a1	−0.113	−0.684	0.052	−0.034	−0.028	34.03	0.83	
	Netz+43a6	−0.115	−0.575	0.051	−0.032	−0.028	34.02	0.67	

Table 7.2: Limiting Kirkwood-Buff integrals (in nm^3), η values which appear in Eq. 7.2 (in nm^{-3}) and $\left(\frac{\partial \Delta G_b}{\partial c_s}\right)_{p,T}$ (in kJ/mol) for ternary solutions of benzene, water and chloride salts of different cations where different force-fields are used for benzene and ions. The labels of the force-fields read as: force-fields for ions + force-fields for benzene. sim: simulation, exp: experimental.

Lithium shows the weakest ion pairing with chloride in water for the cations studied here, which, assuming no direct interactions with benzene, would make lithium chloride the strongest benzene salting-out agent (because of a lower η than other alkali chlorides). Lithium, however, is positioned close to rubidium in the Hofmeister series of the benzene salting-out coefficient, which, as evidenced by the data in Fig. 7.1 and Fig. 7.5, must be due to direct lithium-benzene correlations. As the benzene-water KBIs are similar for all cations, therefore the unusual property of LiCl as salting-out agent indeed follows from the stronger benzene-LiCl short range

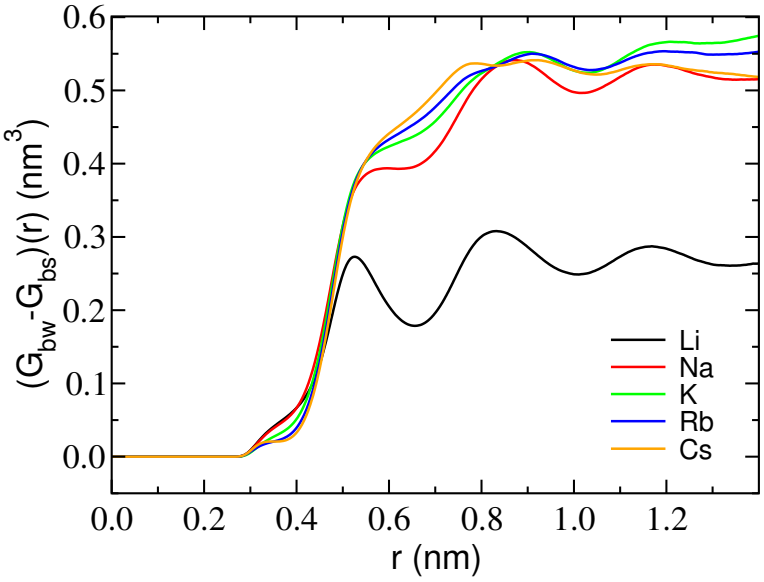


Figure 7.5: Running KB integrals $\mathrm{G}_{ij}(r) = 4\pi \int_0^r [\mathrm{g}_{\mathrm{ij}}(s) - 1]\, s^2 ds$ (with $\mathrm{g}_{\mathrm{ij}}(s)$ being the RDF between i and j) between benzene and water (bw) minus the running KB integrals between benzene and salt (bs), with cations and anions being treated as indistinguishable. Force-fields used– benzene: gromos53a6; ions: Smith.

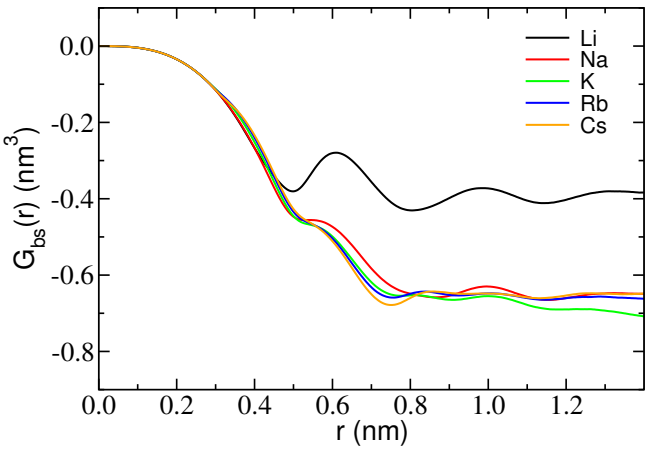


Figure 7.6: Running KB integrals $\mathrm{G}_{ij}(r) = 4\pi \int_0^r [\mathrm{g}_{\mathrm{ij}}(s) - 1]\, s^2 ds$ (with $\mathrm{g}_{\mathrm{ij}}(s)$ being the RDF between i and j) between benzene and salt (with cations and anions being treated as indistinguishable) for 1 m aqueous solutions of alkali chlorides with very dilute concentration of benzene. Force-fields used– benzene: gromos53a6; ions: Smith.

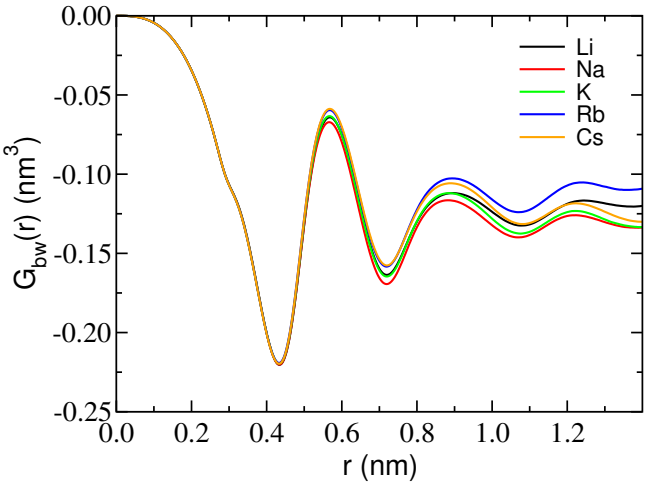


Figure 7.7: Running KB integrals $\mathrm{G}_{ij}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ (with $\mathrm{g_{ij}}(s)$ being the RDF between i and j) between benzene and water for 1 m aqueous solutions of alkali chlorides with very dilute concentration of benzene. Force-fields used– benzene: gromos53a6; ions: Smith.

correlations compared to the other salts. The LiCl-benzene KBI is greater than the corresponding KBIs of all other salts due to water bridging interactions of the hydrated lithium ion with the ring (see Fig. 7.3). We note that the direct affinity between Li^+ and benzene is overestimated in our simulations, due to limitations of the force-field. As mentioned above, the direct interaction mechanism was confirmed by two different Kirkwood-Buff derived Li^+ force field models[32,33] which reproduced experimental KBIs of pure aqueous LiCl solution and also by Dang[35] force-fields. Dang force-field for Li^+ and Cl^- combined with gromos43a1 benzene force-field show comparatively weaker benzene-Li^+ interaction resulting in a higher salting-out coefficient for Li^+, yet underestimated than the experimental result (see table 7.2). A comparison between the benzene-Li^+ RDFs obtained with different force-fields can be found in Fig. 7.11.

LiCl salt shows anomaly in terms of Setchenow constants related to the solvation of pure hydrophobic molecules such as methane or neopentane in salt-water solutions. From the experimental data it is found that the Setchenow constants for methane or neopentane in LiCl-water solutions are lower than that in aqueous NaCl or KCl solutions.[18] On this note, we wanted to examine the effect of the partial charges on the atoms of the benzene molecule on the salting-out phenomena by switching-off the charges on the hydrogen and carbon atoms of benzene molecule and making it a pure hydrophobic molecule. The direct correlations between the benzene molecules (with and without Coulombic charges) and the cations

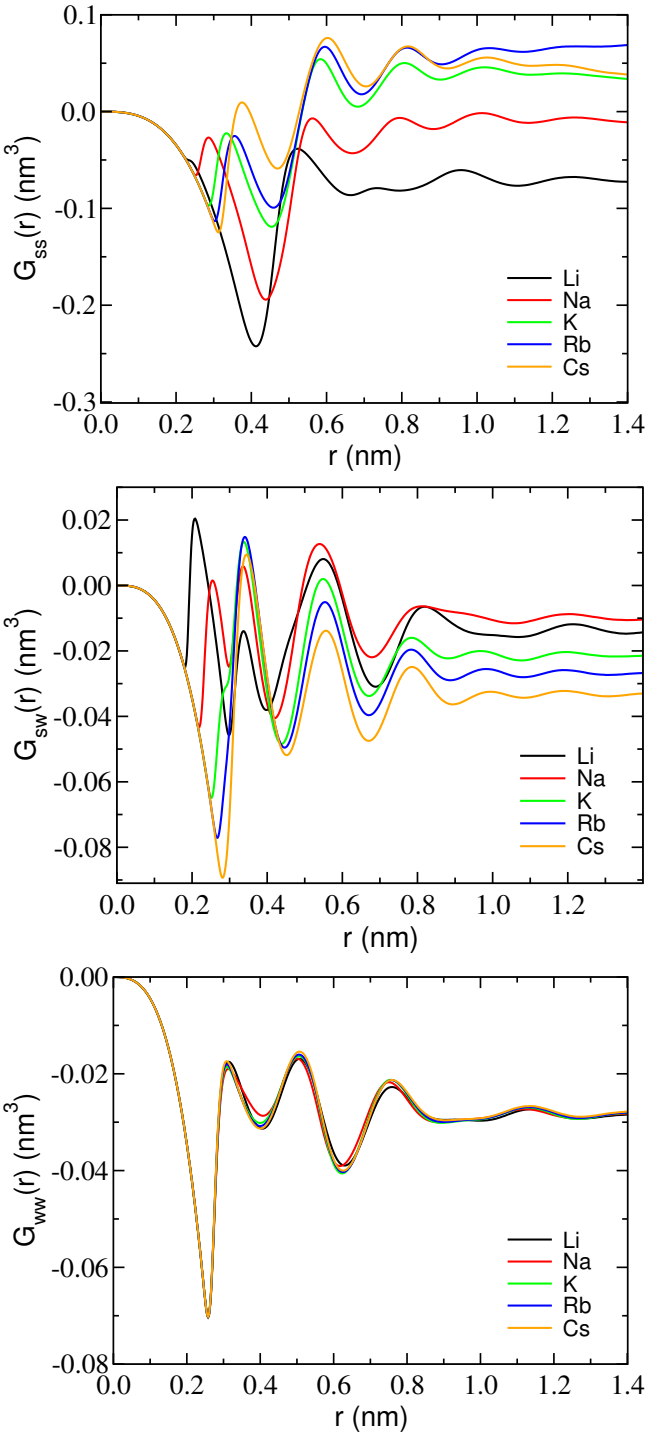


Figure 7.8: Running KB integrals $G_{ij}(r) = 4\pi \int_0^r [g_{ij}(s) - 1]\, s^2 ds$ with $g_{ij}(s)$ being the RDF between solvent components i and j. The upper panel shows the salt-salt KBIs (cation and anions are treated indistinguishable), the middle panel shows the salt-water KBIs and the lower panel the water-water KBIs. Force-fields used– benzene: gromos53a6; ions: Smith.

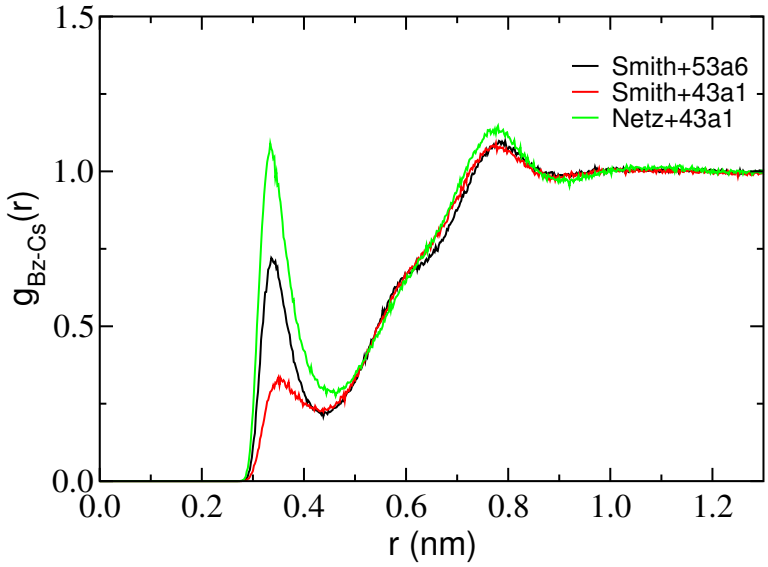


Figure 7.9: Radial distribution functions between benzene and Cs^+ in 1 m aqueous solutions of CsCl with different force-fields for ions (Smith and Netz) and benzene (gromos43a1 and gromos53a6).

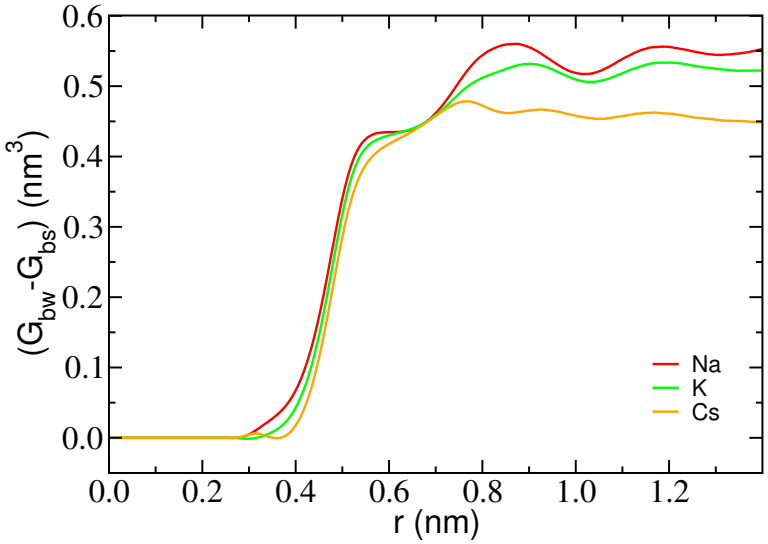


Figure 7.10: Running KB integrals $\mathrm{G}_{ij}(r) = 4\pi \int_0^r [\mathrm{g_{ij}}(s) - 1]\, s^2 ds$ (with $\mathrm{g_{ij}}(s)$ being the RDF between i and j) between benzene and water (bw) minus the running KB integrals between benzene and salt (bs), with cations and anions being treated as indistinguishable. Force-fields used– benzene: gromos43a1; ions: Netz.

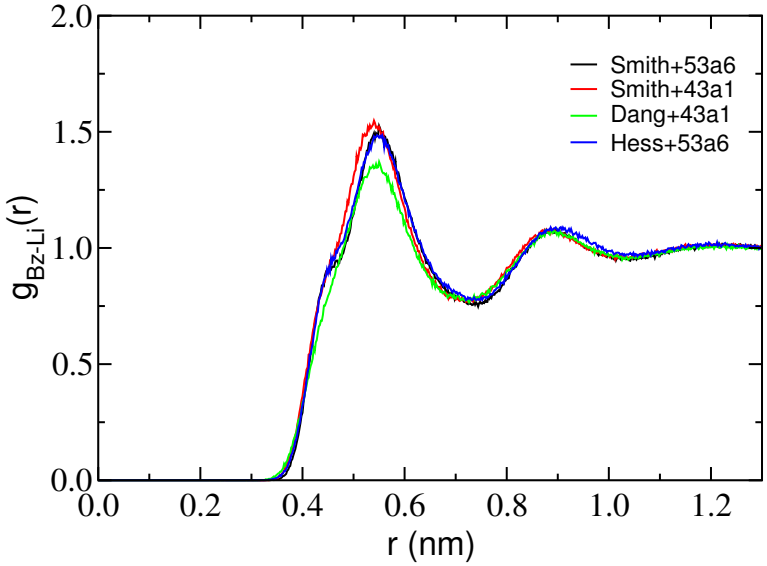


Figure 7.11: Radial distribution functions between benzene and Li^+ in 1 m aqueous solutions of LiCl with different force-fields for ions (Smith, Hess and Dang) and benzene (gromos43a1 and gromos53a6).

(Li^+ and Na^+) are plotted in Fig. 7.12. For these calculations we have used Dang force-field for Li^+ and Cl^-, Netz force-field for Na^+ and Cl^- and gromos43a1 model for benzene. From Fig. 7.12 we find no significant changes in the benzene-Li^+ correlations after switching of the charges on benzene. The KBI between benzene and LiCl salt, G_{bs}, changes from $-0.435\,nm^3$ to $-0.453\,nm^3$ after switching-off the charges on benzene. Clearly switching-off the charges on benzene does not change the solution properties significantly. For Na^+, we find that the first small peak in the benzene-Na^+ RDF is missing when the charges on benzene are switched-off but long-range correlations after 0.6 nm are nearly unaltered. The KBI G_{bs} changes from $-0.650\ nm^3$ to $-0.712\ nm^3$ for benzene with no Coulombic charges. All the other KBIs do not vary significantly or do not vary at all. The corresponding RDFs between benzene and water molecules are also plotted in Fig. 7.13. It is interesting to notice that benzene-water pair-structure also does not change significantly upon switching-off the charges on benzene molecules, except for the missing small first peak at 0.3 nm. Benzene-water KBIs remain almost unaltered. Also it is clear from Fig. 7.13 that the benzene-water RDF and the corresponding KBI do not show any ion-specificity upon changing the cation from lithium to sodium. From these results one can argue that the low partial charges on hydrophobic molecules such as benzene do not play significant role to determine the direct interaction between the hydrophobes and the ions (in terms of the KBIs) and these direct interactions (between the hydrophobes

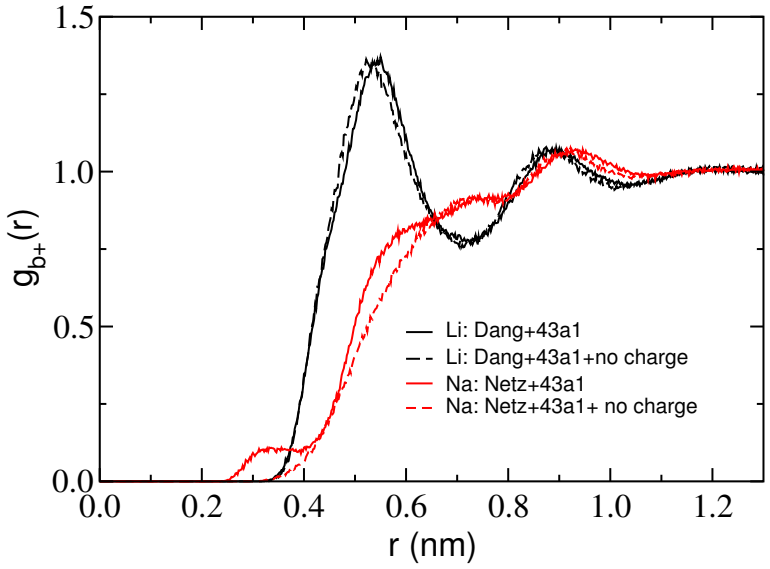


Figure 7.12: Radial distribution functions between benzene and the cations (Li^+ and Na^+) in 1 m aqueous solutions of alkali Chlorides with different force-fields for ions (Dang and Netz) with gromos43a1 benzene (with and without partial charges on the hydrogen and carbon atoms of benzene).

and the cations) which are dominated by the Lennard-Jones interactions seem to be applicable for the pure hydrophobic molecules as methane or neopentane. Thus it also provides a potential explanation for the anomaly of LiCl in terms of salting-out methane or neopentane molecules when compared with NaCl or KCl salt.

7.5 Discussion and Conclusion

Fluctuation theory of solutions has been applied to study the salting-out of benzene from water by dissolved electrolytes. Generally, salting-in or salting-out is determined by the affinities between the solute and the solvent components (the direct correlations), while the affinities between the solvent components (indirect correlations) attenuate or reinforce the magnitude of the observed changes in solute solubility. Here, affinities are quantified by means of Kirkwood-Buff integrals. Eq. 7.2 quantifies the well-established notion that preferential hydration of the solute leads to salting-out, *i.e.* raising the salt concentration reduces the number of water molecules available to hydrate the solute, causing a decrease of its solubility. It is perhaps less well established to what extent the observed salting-out effect depends on differences in affinities between solvent components that give rise to thermodynamic non-ideality of the binary solvent mixture. From our simulations we have

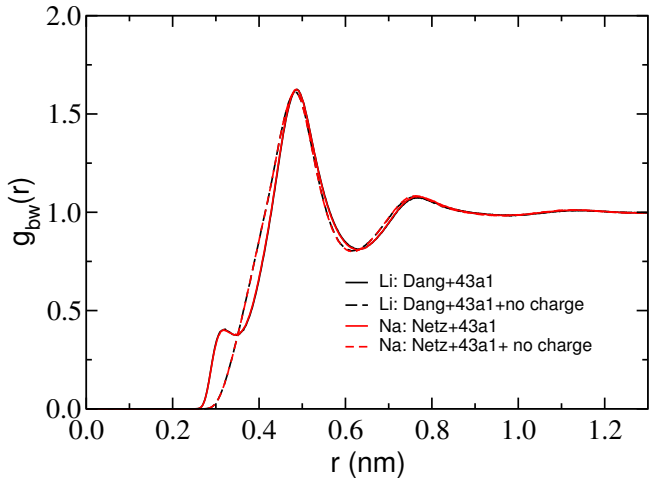


Figure 7.13: Radial distribution functions between benzene and water in 1 m aqueous solutions of lithium and sodium chloride with very dilute concentration of benzene (with and without partial charges). Legends represent the force-fields for ions and benzene respectively with “nocharge" representing the benzene molecules without partial charges on the atoms.

quantitatively analyzed the solvent contribution to the salting-out phenomena for alkali chloride salts in water and found that the difference in the Setchenow constants for different cations can not be explained by the quantity η in Eq. 7.2 which contains the solvent correlations namely salt-salt, salt-water and water-water. Our results show that the quantity $(\mathrm{G_{ww}} + \mathrm{G_{ss}} - 2\mathrm{G_{sw}})$ is small enough to contribute to any significant change in η for different cations as the quantity η is dominated by the number density of water ρ_{w} which is much larger than ρ_{s} and $[\rho_{\mathrm{s}}\rho\mathrm{w}\,(\mathrm{G_{ww}} + \mathrm{G_{ss}} - 2\mathrm{G_{sw}})]$. An ideal solvent mixture is characterized by $(\mathrm{G_{ww}} + \mathrm{G_{ss}} - 2\mathrm{G_{sw}}) = 0$.[25] For the systems studied in this work one can argue that the binary mixtures of alkali chlorides with water do not deviate much from the ideal solutions as the quantities $(\mathrm{G_{ww}} + \mathrm{G_{ss}} - 2\mathrm{G_{sw}})$ are very small and simulation data presented here are consistent with experiments involving binary salt-water mixtures. From Eq. 7.2 a general conclusion can be drawn that for binary mixtures with low cosolvent (as salts in this work) concentrations, which do not show a high non-ideality, the variation in the salting-out phenomena of any solute in these solutions with different cosolvents can not be explained in terms of the binary solvent-cosolvent solution structures and correlations if the variation in the salting-out coefficients for different cosolvents is relatively high. Using different force-fields for the ions and benzene molecules our results strongly indicate that the variation in the salting-out coefficients of the alkali chlorides in water to salt-out benzene can only be determined by the direct interac-

tions between the cations and the benzene molecules. From the distributions of the ions and water molecules around benzene molecules we find that the larger cations, the chaotropes (K^+, Rb^+ and Cs^+), form contact pairs with benzene along the vertical axis to the plane of the benzene-ring and the kosmotropes (Li^+ and Na^+) form water mediated pairs with benzene vertically and horizontally.

Using Kirkwood-Buff integrals if we quantify the direct interactions between benzene and salts we observe that the current Kirkwood-Buff force-fields for the ions are incapable of perfectly reproducing experimental salting-out coefficients for benzene in water though the KBIs for the salt-water binary mixtures are consistent with experiments. Using Netz force-fields[34] for the ions we do observe the experimental salting-out series as $NaCl > KCl > CsCl$ but quantitatively the differences with different cations are less pronounced as in experiments. But for all the simulations we notice that the difference in the Setchenow constants for the salting-out of benzene is caused by the difference in the direct benzene-salt interactions. The direct interactions of anions[6,7,43–46] and cations[47] with proteins and peptide backbones have been discussed in the recent years. Our simulations also show the anomaly of Li^+ when compared to Na^+ or K^+ as Li^+ shows lower salting-out coefficient than Na^+ or K^+ with all the force-fields used but the Setchenow constant for LiCl is always underestimated than the experiments. Using Dang force-fields[35] for LiCl we observe an improvement in the salting-out coefficient of LiCl towards the experimental value through a less-pronounced direct interaction between benzene and Li^+. The interaction between benzene and Li^+ is found to be predominantly Lennard-Jones type in nature and approximately insensitive to the small partial Coulombic charges on the carbon and hydrogen atoms of the benzene molecules. Thus this also provides an evidence for the direct solute-cation interaction mechanism while explaining the anomaly of LiCl for salting-out small pure hydrophobic molecules such as methane or neopentane unlike the indirect mechanism proposed by Elcock $et\,al.$[20] For all of our simulations we do not observe any cation-specific changes in the over-all water-water and benzene-water correlations.

7.6 Acknowledgement

This research was supported by the German Research Foundation (DFG) within the Cluster of Excellence 259 "Smart Interfaces: Understanding and Designing Fluid Boundaries". We thank Francisco Rodríguez-Ropero for useful discussions.

Bibliography

[1] Hofmeister, F. *Arch. Exp. Pathol. Pharmakol.* **1888**, *24*, 247.

[2] Collins, K. D. *Proc. Natl. Acad. Sci. USA* **1995**, *92* 5553.

[3] Baldwin, R. L.; *Biophys. J.* **1996**, *71*, 2056.

[4] Cacace, M. G.; Landau, E. M.; Ramsden, J. J. *Q. Rev. Biophys.* **1997**, *30*, 241.

[5] Kunz, W.; Lo Nostro, P.; Ninham, B. W. *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 1.

[6] Zhang, Y. J.; Cremer, P. S. *Curr. Opin. Chem. Biol.* **2006**, *10*, 658.

[7] Zhang, Y. J.; Cremer, P. S. *Annu. Rev. Phys. Chem.* **2010**, *61*, 63.

[8] Lo Nostro, P.; Ninham, B. W. *Chem. Rev.* **2012**, *112*, 2286.

[9] Ion specific Hofmeister Effects *Faraday Disc.* **2013**, *160* .

[10] Omta, A. W.; Kropman, M. F.; Woutersen, S.; Bakker, H. J. *Science* **2003**, *301*, 347.

[11] Gurau, M. C.; Lim, S. M.; Castellana, E. T.; Albertorio, F.; Kataoka, S.; Cremer, P. S. *J. Am. Chem. Soc.* **2004**, 126, 10522.

[12] Qvist, J.; Halle, B. *J. Am. Chem. Soc.* **2008**, *130*, 10345.

[13] Zhang, Y.; Furyk, S.; Bergbreiter, D. E.; Cremer, P. S. *J. Am. Chem. Soc.* **2005**, *127*, 14505.

[14] Zangi, R.; Hagen, M.; Berne, B. J. *J. Am. Chem. Soc.* **2007**, *129*, 4678.

[15] Samoilov, O. Y. *Discuss. Faraday Soc.* **1957**, *24*, 141.

[16] von Hippel, P. H.; Schleich, T. *Acc. Chem. Res.* **1969**, *2*, 257.

[17] Long, F. A.; McDevit, W. F. *Chem. Rev.* **1952**, *51*, 119.

[18] Weisenberger, S.; Schumpe, A. *AIChE J.* **1996**, *42*, 298.

[19] Setchenow, J. *Z. Phys. Chem.* **1889**, *4*, 117.

[20] Thomas, A. S.; Elcock, A. H. *J. Am. Chem. Soc.* **2007**, *129*, 14887.

[21] Graziano, G. *J. Chem. Eng. Data* **2009**, *54*, 464.

[22] Athawale, M. V.; Sarupriya, S.; Garde, S. *J. Phys. Chem. B* **2008**, *112*, 5661.

[23] Tomé, L. I. N.; Varanda, F. A.; Freire, M. G.; Marrucho, I. M.; Coutinho, J. A. P. *J. Phys. Chem. B* **2009**, *113*, 2815.

[24] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

[25] Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.

[26] Ganguly, P.; van der Vegt, N. F. A. *J. Chem. Theory Comput.* **2013**, *9*, 1347.

[27] Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, *7*, 306.

[28] van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Hochschulverlag AG an der ETH Zürich (1996).

[29] Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656.

[30] Berendsen H. J. C.; Grigera J. R.; Straatsma T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

[31] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *119*, 11342.

[32] Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. *J. Chem. Theory Comput.* **2011**, *7*, 1369.

[33] Hess, B.; van der Vegt, N. F. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 13296.

[34] Fyta, M.; Netz, R. R. *J. Chem. Phys.* 2012, *136*, 124103.

[35] Dang, L. X. *J. Am. Chem. Soc.* **1995**, *117*, 6954.

[36] Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

[37] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

[38] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

[39] Essmann U.; Perera L.; Berkowitz M. L.; Darden T.; Lee H.; Pedersen L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

[40] Fennell, C. J.; Bizjak A.; Vlachy, V.; Dill, K. A. *J. Phys. Chem. B* **2009**, 113, 6782.

[41] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J.; *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, 1981; pp 331-342.

[42] Schravendijk, P.; van der Vegt, N. F. A. *J. Chem. Theory Comput.* **2005**, *1*, 643.

[43] Tomé, L. I. N.; Jorge M.; Gomes, J. R. B.; Coutinho, J. A. P. *J. Phys. Chem. B* **2010**, 114, 16450.

[44] Heyda, J.; Vincent J. C.; Tobias, D. J.; Dzubiella, J.; Jungwirth, P. *J. Phys. Chem. B* **2010**, 114, 1213.

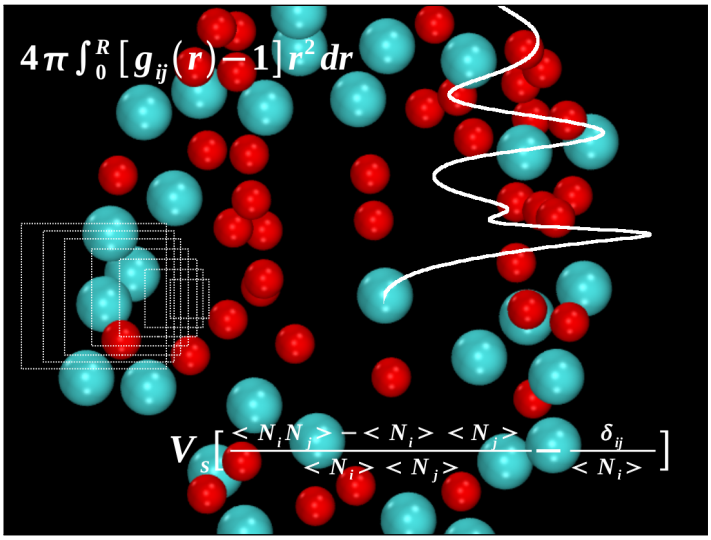
[45] Algaer, E. A.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2011**, 115, 13781.

[46] Rembert, K. B.; Paterova J.; Heyda, J.; Hilty, C.; Jungwirth, P.; Cremer, P. S. *J. Am. Chem. Soc.* **2012**, 134, 10039.

[47] Okur, H. I.; Kherb, J.; Cremer, P. S. *J. Am. Chem. Soc.* **2013**, 135, 5062.

8 Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations

We discuss two methods for calculating Kirkwood-Buff integrals (KBIs) of aqueous cosolvent solutions from molecular simulations. The first method is based on computing running integrals over radial distribution functions obtained from NVT or NpT simulations. The second, more recent method, originally introduced by Schnell *et al.* (*J. Phys. Chem. B* **2011**, *115*, 10911), obtains the KBIs from direct analysis of particle number fluctuations in small, open subvolumes embedded in a larger reservoir as provided by the NVT (NpT) simulation cell. The thermodynamic limit is taken in the first method by using the plateau-values of the running KBIs for large distances, while in the second method an analytical finite-size scaling relation is applied to the KBIs of subvolumes of variable size. We find that direct analysis of particle number fluctuations at small scales provides more precise estimates of KBIs for methanol-water and urea-water solutions. Converged KBIs could however not be obtained from nanosecond time scale molecular dynamics simulations with either of the two methods. Based on 0.1 μs simulation trajectories of small and large system sizes time-converged KBIs were obtained with both methods. The running integral method suffers however from stronger finite-size artifacts than the sub-box method, also when empirical finite-size tail corrections are applied to the radial distribution functions.



8.1 Introduction

The Kirkwood-Buff (KB) theory[1] proposed in 1951 relates thermodynamic quantities of stable solution mixtures to the microscopic liquid structure. The theory is derived in the grand-canonical (μVT) ensemble with any number of components of any type and defines the so-called Kirkwood-Buff integrals (KBIs), which are the integrals of the radial distribution functions (RDFs) over volume. The KBI between mixture components i and j is defined as

$$G_{ij} = 4\pi \int_0^\infty \left[\mathrm{g}_{ij}^{\mu VT}(r) - 1 \right] r^2 dr, \tag{8.1}$$

where $\mathrm{g}_{ij}^{\mu VT}(r)$ is the RDF between components i and j. These integrals can be related to thermodynamic quantities like the isothermal compressibility, partial molar volumes, and derivatives of the chemical potentials or activity coefficients of solution components with solution composition.[2] Physically, KBIs provide a measure of the mutual affinity between solution components; *i.e.* the quantity $\rho_j G_{ij}$, with ρ_j being the number density of molecules j, is the change in the average number of molecules j in a spherical region of radius R caused by placing a molecule i at the center of the region. Here R is defined as $\mathrm{g}_{ij}(r \geq R) = 1$ and typically corresponds to a distance between 1.0 and 2.0 nm, depending on the system under study. Hence KB theory provides a link between *local* (< 2.0 nm) properties (G_{ij}s) and global, thermodynamic properties. KB theory is exact, does not assume pair-wise additivity of the potentials and provides a powerful and computationally straightforward route to obtain thermodynamic properties from the RDFs obtained with molecular simulations. Using the inversion of the Kirkwood-Buff theory[3] KBIs can moreover be obtained experimentally. This provides a means for parameterizing atomistic force-fields based on KB theory, early applications of which included aqueous cosolvent mixtures[4–9], aqueous amide mixtures[10], to name a few. Recently, KB theory has also been used to parametrize a coarse-grained urea-water force field.[11]

Although KBIs (Eq. 8.1) should in principle be obtained from calculations in open systems, they are usually obtained from closed-boundary NpT or NVT simulations.[4–12] This is justified, provided that the density fluctuations are local and the simulation-box size is significantly bigger than any of the correlation lengths

which are typically < 2.0 nm for aqueous solutions far away from critical points. If this condition is met, the KBI is a local quantity and the solution thermodynamics can be related to the solution structure in NpT and NVT simulations. For closed boundary systems, KBIs (G_{ij}) calculated (Eq. 8.1) by integrating over the overall box volume take the values $-1/\rho_j$ ($i = j$) or 0 ($i \neq j$). Clearly, these are not the required results. However, given the condition that the density fluctuations are local, we expect the integral to asymptotically approach the correct KBI for integration volumes smaller than the box size. The running-KBI (RKBI), defined as

$$G_{ij}(r) = 4\pi \int_0^r \left[\mathrm{g}_{ij}^{NpT}(r') - 1 \right] r'^2 dr', \tag{8.2}$$

thus approaches a plateau value that corresponds to the thermodynamic limit in Eq. 8.1 at a distance r, which is small compared to the box size. In practice, KBIs calculated this way are inaccurate due to the following reasons: (1) Poor convergence of the RDFs at longer distances as it takes long for the particles to move over large distances and consequently the tails of the RDFs suffer from poor statistics. This error is further weighted with r^2 while calculating the KBIs, (2) RKBIs calculated in closed systems do not show the correct asymptotic behavior because the RDFs go to a limit which is not one.[2,13–15] The second issue has not been considered in several earlier works considering the analysis of RKBIs[4–11] where the authors obtained the KBIs by taking averages of the RKBIs in a finite range of the radial separations. Perera and Sokolić rescaled the pair correlation function in such a way that it approaches one asymptotically at half of the simulation box length[14] and, in a later work, used an alternative approach.[15] Another method of rescaling the RDFs was used by Hess and van der Vegt where the RDFs were renormalized with the actual particle densities in the bulk at large distance.[16] Christensen *et al.*[17] and Wedberg *et al.*[18] used parametric functions to fit RDFs and further extrapolated them to long-range.

Schnell *et al.* proposed an alternative method to calculate the thermodynamic limiting values of the KBIs from simulations of small boxes.[19] In their earlier work they derived[20] expressions for thermodynamic quantities like the thermodynamic correction factor (that appears in Fick diffusion coefficients) and molar enthalpy of small μVT systems using the formalism proposed by Hill.[21] Based on this formalism they showed that the inverse of the thermodynamic correction factor and molar enthalpy of small μVT systems scale linearly with the inverse of the linear dimen-

sion of the system. In their later work[19] they calculated KBIs of small non-periodic systems embedded in a larger periodic system and showed that the KBIs also scale linearly with the inverse of the linear dimension of the small systems as

$$G_{ij}(L_s) = G_{ij} + \frac{A'}{L_s} \tag{8.3}$$

where $G_{ij}(L_s)$ is the KBI of the small embedded system (from this point on referred to as SKBI), G_{ij} is the SKBI in the thermodynamic limit for $L_s \rightarrow \infty$ and L_s is the linear size of the cubic small system. These small cubic boxes (referred to as sub-boxes from now on) can be thought to represent grand-canonical systems where the bigger simulation box acts as a large particle bath or reservoir with which the smaller sub-boxes can exchange energy and particles. This picture can be maintained provided that the size (L_s) of the sub-box is sufficiently smaller than the larger (NVT or NpT) simulation box with linear size L. The large periodic simulation boxes and small non-periodic sub-boxes are shown schematically in Figure 8.1. The SKBIs were calculated from the particle number fluctuation as[19]

$$G_{ij}(L_s) = V_s \left[\frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_i \rangle} \right] \tag{8.4}$$

where N_i is the number of particles of type i within the sub-box, V_s is the volume of the sub-box, and δ_{ij} is Kronecker delta. $\langle .. \rangle$ denotes the grand-canonical ensemble average. Calculations of $G_{ij}(L_s)$ for various sub-box sizes L_s allow to obtain the thermodynamic limiting value G_{ij} by means of Eq. 8.3.

In the present study we simulate urea-water and methanol-water solutions with different system sizes using Kirkwood-Buff-derived atomistic force fields.[5,7] We calculate KBIs using Schnell's sub-box method[19] with cubic and spherical sub-boxes and compare the results with KBIs obtained from the RKBIs (Eq. 8.2) in the limit of large r. As the force fields were originally developed using finite-sized systems, the analysis of finite-size effects provides additional information on the accuracy of the models. We show that KBIs obtained by taking the average values of the, usually oscillating, RKBIs in an arbitrary distance range contain relatively large fluctuations, since the KBIs depend on the range of distances used for averaging as well as on system size. We also study a finite-size correction to the RDFs and its effect on the KBIs calculated with different system sizes. Convergence issues of RKBIs and SKBIs, ob-

tained from finite-time molecular dynamics simulations, are furthermore examined.

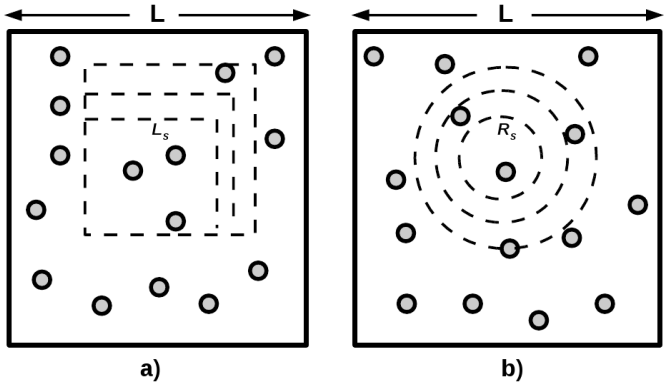


Figure 8.1: Cubic (a) and spherical (b) subsystems (dashed) embedded in a periodic simulation box of length L. L_s and R_s are the variable linear dimension and radius of the cubic and spherical subsystems, respectively. These open subsystems are non-periodic, the periodic simulation box with the larger dimension L provides a particle-bath for the small subsystems.

8.2 Computational Details

All-atomistic simulations were performed with the GROMACS molecular dynamics package (version 4.0).[22] The nonbonded parameters for methanol and urea were taken from Kirkwood-Buff-theory-derived force-fields[5,7] as mentioned earlier. The SPC/E water model was used for both the systems.[23] NpT simulations were carried out at a temperature of 300 K and a pressure of 1 bar. The temperature and pressure were kept constant using Nose-Hoover thermostat[24,25] and Parrinello-Rahman barostat,[26] respectively. Electrostatic interactions were evaluated using the particle mesh Ewald (PME) method.[27] The cut-off of non-bonded interactions was taken to be 1 nm.

8.3 Results and Discussions

8.3.1 Convergence of KBIs with simulation time

Urea-water and methanol-water mixtures were simulated using large systems (approximately 10000 to 12000 molecules in total) as well as small systems (approximately 2000 to 2500 molecules in total). 100-ns-long trajectories were accumulated

with an integration time-step of 2 fs. The RKBIs and SKBIs were obtained from 100 ns simulations. In particular the long range part of the RKBI equilibrates very slowly as shown in Figure 8.2 and Figure 8.3 for the methanol-water and urea-water systems, respectively. Clearly, the limiting values of the RKBIs at distances larger than 1.0 nm cannot be obtained with reasonable accuracy from short (*i.e.* 5 ns) production runs. We note that equally long (100 ns) simulations for the same urea-water system (as in Figure 8.3) are required to obtain converged SKBIs as shown by the data presented in Figure 8.4.

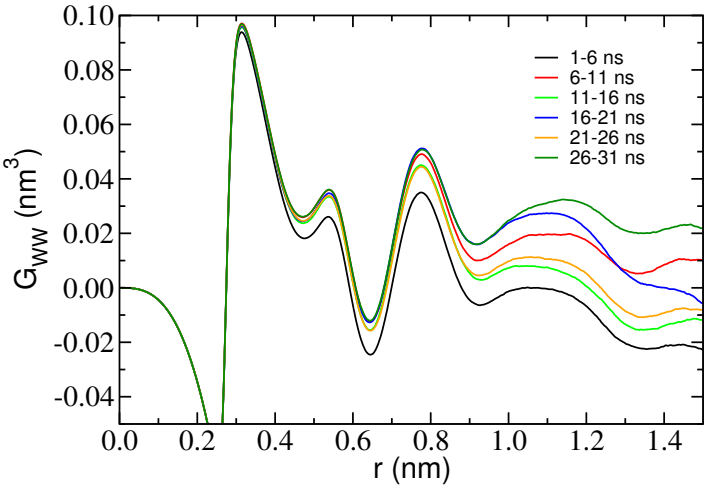


Figure 8.2: Water-water RKBIs for a methanol-water mixture (70 mol% methanol) calculated from simulations of a NpT system with 2000 molecules. RKBIs are presented for different sampling time intervals of 5 ns taken from a 100 ns trajectory.

The strong run length dependence shown in the data probably reflects the characteristic micro-heterogeneous nature of aqueous solutions[28–31] with a correspondingly slow domain like dynamics. Urea self-aggregation and correspondingly slow equilibration of concentrated aqueous urea solutions has previously been reported.[32,33] Neutron diffraction experiments and molecular simulations of mixtures of methanol and water showed that these systems exhibit extended structures, which, within a certain concentration region appear to form separate, percolating networks.[34] In contrast to aqueous solutions of nonelectrolytes, oscillations of RKBIs of aqueous electrolyte solutions (≈ 1 m salt concentration) typically cease when r approaches 1 nm.[12,35]

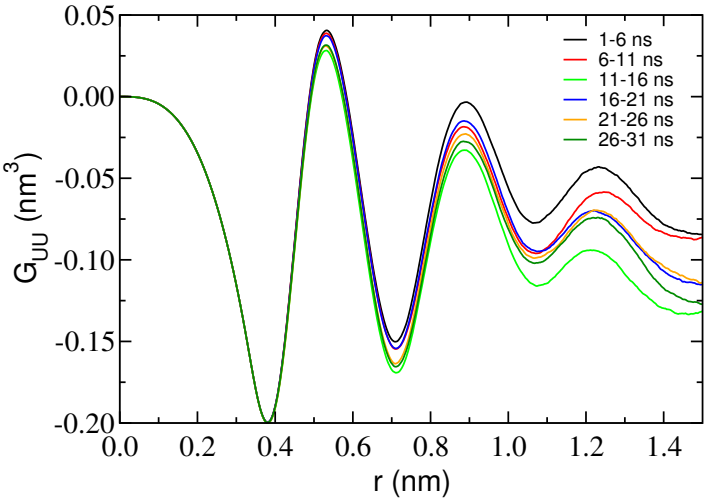


Figure 8.3: Urea-urea RKBIs for a urea-water mixture (6 m urea) calculated from simulations of a NpT system with 2000 water molecules. RKBIs are presented for different sampling time intervals of 5 ns taken from a 100 ns trajectory.

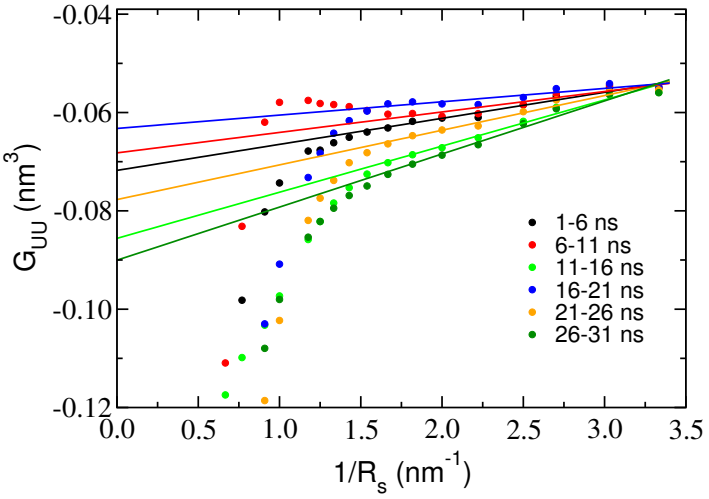


Figure 8.4: Urea-urea SKBIs obtained with the sub-box method for a urea-water mixture (6 m urea) calculated from simulations of a NpT system with 2000 water molecules. The particle number fluctuations have been calculated (Eq. 8.4) with different sampling time intervals of 5 ns taken from a 100 ns trajectory. The solid lines are fitted to the linear parts of the data. The linear extrapolations ($R_s \to \infty$) provide the thermodynamic limiting KBIs.

8.3.2 Urea-water KBIs from large system sizes

For urea-water large cubic boxes (7.5-8.0 nm) with periodic boundary conditions were simulated. The urea concentration was varied from 6 m (4.7 M) to 12 m

(7.7 M). The urea-urea, urea-water and water-water KBIs were calculated using two methods: (1) from the direct integration of the respective RDFs (Eq. 8.2) and (2) from the particle number fluctuations (Eq. 8.4). In order to get the KBIs from particle number fluctuations we inserted non-periodic sub-boxes into the simulation box and counted the number of the urea and water molecules within the sub-boxes. Statistical averages were taken by inserting 5000 sub-boxes per frame at random positions within the simulation box along with a further averaging done using 15000 frames which span 90-ns-long trajectories. The shape of the embedded sub-boxes was chosen to be cubic and the SKBIs were calculated for sub-boxes with varying linear size L_s. Figure 8.5 shows the SKBIs between urea-urea (G_{UU}), urea-water (G_{UW}) and water-water (G_{WW}) as functions of the inverse of the linear dimension (L_s) of the cubic sub-boxes for different urea concentrations. It is interesting to see that we find a regime corresponding to the sub-box length-scale of approximately 0.7 nm to 1.5 nm where the SKBIs vary linearly with the inverse of the sub-box lengths. This regime is used to obtain the limiting KBIs by linear extrapolation into the thermodynamic limit $1/L_s \to 0$ according to Eq. 8.3. The limiting KBIs derived by this procedure are reported in Table 8.1. The limiting KBIs obtained from the direct integration of the RDFs between urea-urea, urea-water and water-water are also shown in Figure 8.5 (represented by dashed horizontal lines). These values were calculated by averaging the RKBIs (Eq. 8.2) in a distance range between r =1.0 and 1.4 nm. We can see that the KBIs obtained with the two methods are in reasonable agreement although the values obtained from the RKBIs may fluctuate 10–15 $\%$ if the spatial region used for averaging the RKBIs changes. Table 8.1 shows the variation of the KBIs obtained from RKBIs by carrying out the averaging in three different regions, namely from 0.8 to 1.2 nm, from 1.0 to 1.4 nm and from 1.1 to 1.5 nm. For illustration, the gray curves in Figure 8.5 show the RKBIs at 6 m urea-concentration. We see that the RKBIs are still oscillating even after 1.0–1.5 nm which leads to smaller precision (error 10–15 $\%$) of the KBIs obtained with the integration method compared to the limiting KBIs ($L_s \to \infty$) obtained from SKBI calculations. The observed, drifting asymptotes of the RKBIs shown in Figure 8.5 are due to finite system size effects and can be alleviated by applying corrections to the tails of the RDFs as discussed in the introduction. In particular for smaller systems (2000 molecules, see below), tail corrections are required since direct integration of the uncorrected RDFs yields largely oscillating and drifting RKBIs. This aspect is discussed in greater detail later on.

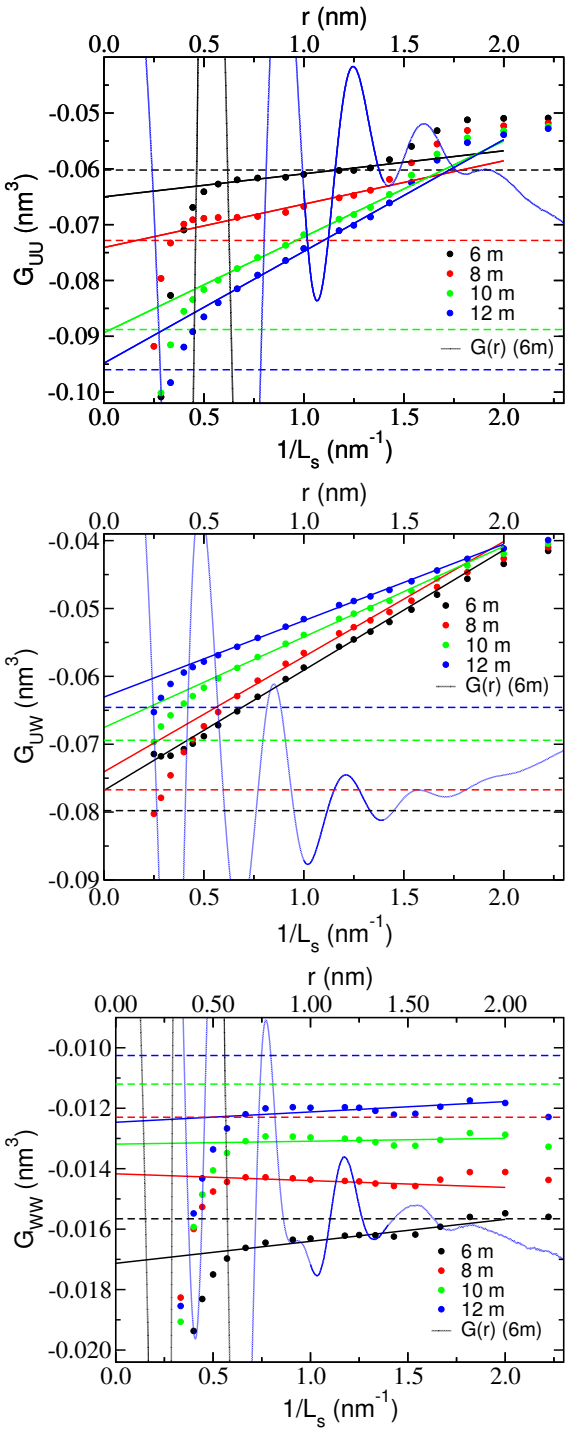


Figure 8.5: Urea-urea (upper panel), urea-water (middle panel) and water-water (lower panel) SKBIs calculated from the particle number fluctuations of cubic sub-boxes presented as functions of the inverse of the sub-box lengths for urea-water mixtures (big-box, 11111 water molecules) with different urea concentrations (molality). Solid straight lines are fitted to the linear regime of the plots. Horizontal dashed lines denote the limiting KBIs obtained from direct integration of the corresponding RDFs. Gray lines are the RKBIs at 6 m urea-concentration. The averaging interval used to obtain the limiting KBIs is darkened on the curve.

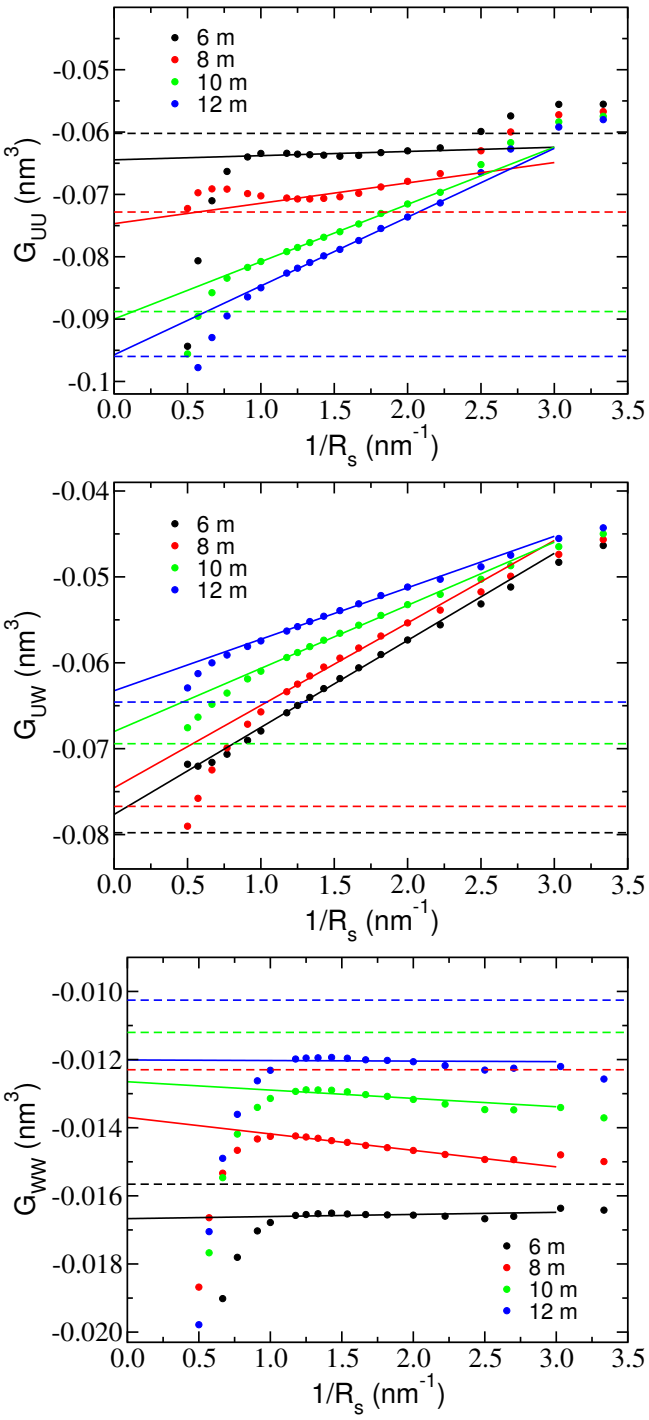


Figure 8.6: Urea-urea (upper panel), urea-water (middle panel) and water-water (lower panel) SKBIs calculated from the particle number fluctuations of spherical sub-boxes as functions of the inverse of the sub-box radii for urea-water mixtures (big-box, 11111 water molecules) with different urea concentrations (molality). Solid straight lines are fitted to the linear regime of the plots. Horizontal dashed lines denote the limiting KBIs obtained from direct integration of the corresponding RDFs.

Urea molality	G_{ij}	Integration of RDF						Cubic sub-box	Spherical sub-box	
		bigbox			smallbox			bigbox	bigbox	smallbox
		$0.8-1.2$	$1.0-1.4$	$1.1-1.5$	$0.8-1.2$	$1.0-1.4$	$1.1-1.5$			
6	G_{UU}	-0.052	-0.060	-0.056	-0.064	-0.082	-0.085	-0.065	-0.064	-0.082
	G_{UW}	-0.076	-0.080	-0.078	-0.073	-0.074	-0.070	-0.077	-0.078	-0.073
	G_{WW}	-0.016	-0.016	-0.015	-0.016	-0.017	-0.017	-0.017	-0.017	-0.017
8	G_{UU}	-0.067	-0.073	-0.068	-0.081	-0.096	-0.098	-0.074	-0.075	-0.092
	G_{UW}	-0.072	-0.077	-0.075	-0.067	-0.068	-0.065	-0.074	-0.075	-0.069
	G_{WW}	-0.013	-0.013	-0.012	-0.014	-0.015	-0.015	-0.014	-0.014	-0.015
10	G_{UU}	-0.080	-0.089	-0.085	-0.088	-0.104	-0.106	-0.089	-0.090	-0.096
	G_{UW}	-0.072	-0.069	-0.075	-0.062	-0.063	-0.059	-0.068	-0.068	-0.065
	G_{WW}	-0.011	-0.011	-0.011	-0.013	-0.014	-0.014	-0.013	-0.013	-0.013
12	G_{UU}	-0.087	-0.096	-0.093	-0.089	-0.102	-0.100	-0.095	-0.096	-0.099
	G_{UW}	-0.061	-0.065	-0.062	-0.059	-0.062	-0.059	-0.063	-0.063	-0.062
	G_{WW}	-0.010	-0.010	-0.010	-0.011	-0.012	-0.012	-0.012	-0.012	-0.011

Table 8.1: Limiting KBIs for urea-water mixtures (in units nm^3) obtained from direct integration of RDFs (Eq. 8.2) and from particle number fluctuations (Eq. 8.4) using cubic sub-boxes and spherical sub-boxes by extrapolating to the thermodynamic limit (Eq. 8.3). Data are shown for big and small system sizes. For direct integration of RDFs three different regions for averaging were chosen, namely from 0.8 to 1.2 nm (0.8–1.2), from 1.0 to 1.4 nm (1.0–1.4) and from 1.1 to 1.5 nm (1.1–1.5).

The SKBIs start to deviate from linearity with $1/L_s$ when the length of the sub-boxes is larger than $\approx$ 1.5 nm. This happens because, for the bigger sub-boxes, the simulation box (L =7.5 – 8.0 nm) no longer acts as a sufficiently large particle reservoir, *i.e.* the sub-boxes can no longer be considered as grand-canonical sub-systems. This deviation was also seen in previous works.[19,20] For smaller sub-boxes ($L_s < \approx$ 0.7 nm) where the size of the boxes can be compared to the molecular dimension, the KBIs cannot be calculated correctly from the particle number fluctuations because the particle-count within the sub-boxes would depend on the packing of the molecules at very small length scales and on the shape of the sub-boxes itself. We further examined the KBIs obtained using spherical sub-boxes with same averaging criteria as for cubic sub-boxes. The SKBIs calculated from particle number fluctuation within the spherical sub-boxes are shown in Figure 8.6 as functions of the inverse of the radii (R_s) of the sub-boxes. Again we observe a linear regime in the plots between $R_s \approx$ 0.3 to 1.4 nm which can readily be extrapolated ($1/R_s \to 0$) to obtain the KBIs in the thermodynamic limit. These limiting values are also listed in Table 8.1. For both shapes of the sub-boxes, *i.e.* cubical and spherical, we find a comparatively smaller linear regime for urea-urea SKBIs when plotted with the inverse of the linear dimension of the sub-boxes ($L_s \approx$ 0.8 to 1.3 nm and $R_s \approx$ 0.45 to 1.2 nm) as they suffer from poorer statistics than urea-water and water-water SKBIs. If we compare the SKBIs obtained from cubic and spherical sub-boxes we see that the extension of the linear regime is larger with spherical sub-boxes than with cubic sub-boxes. Also for

urea-water and water-water SKBIs spherical sub-boxes do a slightly better job than cubic sub-boxes at the smaller length scales which can be expected from the spherical symmetry of the hydration shells around urea or water molecules, so spherical sub-boxes provide better packing of the particles.

8.3.3 Methanol-water KBIs from large system sizes

A series of methanol-water systems were also studied with methanol mole fractions ranging between 0.1 to 0.9. Each system consisted of 10000 molecules (methanol+water) with cubic simulation boxes of length 7-9 nm having periodic boundaries. KBIs between methanol-methanol (G_{MM}), methanol-water (G_{MW}) and water-water (G_{WW}) were calculated from the particle number fluctuations within the sub-boxes of varying linear dimension embedded in the simulation box. Figure 8.7 shows the resulting SKBIs presented versus the inverse spherical sub-box radius. The SKBIs again follow Eq. 8.3 for a regime $R_s \approx 0.3$ to 1.5 nm and their

Methanol %	**G_{ij}**	**Integration of RDF**						**Cubic sub-box**	**Spherical sub-box**	
		bigbox			smallbox			bigbox	bigbox	smallbox
		$0.8-1.2$	$1.0-1.4$	$1.1-1.5$	$0.8-1.2$	$1.0-1.4$	$1.1-1.5$			
10%	G_{MM}	-0.057	-0.064	-0.062	-0.108	-0.100	-0.078	-0.066	-0.063	-0.078
	G_{MW}	-0.063	-0.064	-0.063	-0.051	-0.055	-0.057	-0.062	-0.064	-0.060
	G_{WW}	-0.020	-0.020	-0.020	-0.023	-0.022	-0.021	-0.021	-0.021	-0.021
30%	G_{MM}	-0.068	-0.074	-0.072	-0.076	-0.076	-0.069	-0.073	-0.073	-0.074
	G_{MW}	-0.059	-0.061	-0.059	-0.054	-0.057	-0.057	-0.058	-0.059	-0.057
	G_{WW}	0.005	0.004	0.004	0.000	0.001	0.003	0.001	0.002	0.001
50%	G_{MM}	-0.072	-0.077	-0.076	-0.078	-0.079	-0.073	-0.075	-0.076	-0.077
	G_{MW}	-0.045	-0.048	-0.046	-0.043	-0.045	-0.043	-0.045	-0.046	-0.045
	G_{WW}	0.029	0.028	0.028	0.019	0.021	0.024	0.022	0.025	0.023
70%	G_{MM}	-0.070	-0.075	-0.074	-0.074	-0.076	-0.070	-0.073	-0.074	-0.073
	G_{MW}	-0.027	-0.031	-0.029	-0.026	-0.028	-0.025	-0.031	-0.030	-0.029
	G_{WW}	0.030	0.030	0.030	0.007	0.012	0.019	0.024	0.027	0.021
90%	G_{MM}	-0.066	-0.072	-0.070	-0.069	-0.071	-0.065	-0.069	-0.069	-0.068
	G_{MW}	-0.014	-0.019	-0.017	-0.016	-0.019	-0.014	-0.020	-0.019	-0.020
	G_{WW}	-0.003	-0.007	-0.010	-0.016	-0.010	-0.000	-0.002	-0.008	-0.002

Table 8.2: Limiting KBIs for methanol-water mixtures (in units nm^3) obtained from direct integration of RDFs (Eq. 8.2) and from particle number fluctuations (Eq. 8.4) using cubic sub-boxes and spherical sub-boxes by extrapolating to the thermodynamic limit (Eq. 8.3). Data are shown for big and small system sizes. For direct integration of RDFs three different regions for averaging were chosen, namely from 0.8 to 1.2 nm (0.8–1.2), from 1.0 to 1.4 nm (1.0–1.4) and from 1.1 to 1.5 nm (1.1–1.5).

values in the thermodynamic limit are reported in Table 8.2. The limiting KBIs calculated using cubic sub-boxes are also listed in Table 8.2. For cubic sub-boxes we found the linear regime to be $L_s \approx 0.5$ to 1.75 nm (data not shown). The SKBIs which suffer from poorer statistics, *i.e.* methanol-methanol at lower methanol concentrations (10%, 30%) and water-water at higher methanol concentrations (70%,

90%), show a comparatively narrower linear regime ($R_s \approx$ 0.45 to 1.2 nm for spherical sub-boxes and $L_s \approx$ 0.6 to 1.3 nm for cubic sub-boxes). KBIs obtained from the direct integration of the corresponding RDFs are also presented in Table 8.2 (with three averaging regions, from 0.8 to 1.2 nm, from 1.0 to 1.4 nm and from 1.1 to 1.5 nm) and in Figure 8.7 (see horizontal dashed lines) where the average values were taken between 1.0 and 1.4 nm.

8.3.4 KBIs with varying simulation box size

8.3.4.1 Effect of RDF tail corrections on RKBIs of small and large methanol-water systems

KBIs evaluated based on simulations of finite-sized boxes suffer from finite-size effects as discussed in the introduction. To illustrate this problem we simulated methanol-water mixtures at 50% methanol concentration with varying box sizes, namely with 10000, 5000 and 2000 molecules in total with simulation box-lengths of 7.9, 6.3 and 4.6 nm, respectively. The RKBIs between methanol-methanol, methanol-water and water-water are shown in Figure 8.8 as a function of the radial distance (see the solid curves). It is clear that we hardly find a distance range with a well-defined plateau needed to estimate the KBIs in the thermodynamic limit. For the smaller systems the RKBIs deviate significantly from those of the bigger systems after $\approx$ 1.0 nm, particularly the methanol-water and water-water RKBIs. Given this problem we wanted to test how the KBIs obtained from particle number fluctuations within the sub-boxes depend on the overall simulation box-size. The respective SKBIs calculated with spherical sub-boxes are also shown in Figure 8.8 presented versus the inverse of the sub-box radius (see solid dots) and the fits of the linear regime are shown as dashed straight lines. The SKBIs and extrapolated KBIs do not vary much with system size (see green, red and black dashed straight lines) while the RKBIs do not converge to any plateau, rendering the integration approach more erroneous in predicting the limiting KBIs. Hence, thermodynamic quantities obtained by analyzing particle number fluctuations in small subvolumes exhibit reasonable accuracy, even if small overall (NVT or NpT) systems are simulated where direct integration of RDFs seems not to obtain a steady limiting KBI.

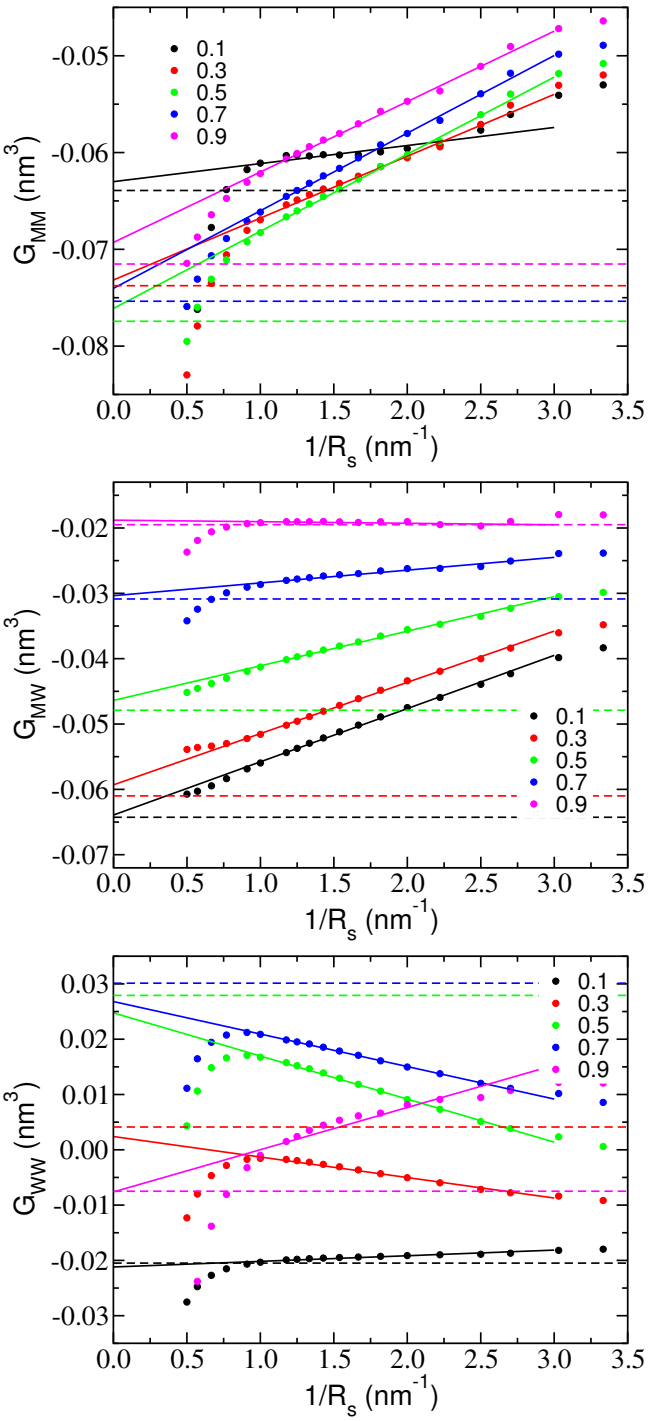


Figure 8.7: Methanol-methanol (upper panel), methanol-water (middle panel) and water-water (lower panel) SKBIs calculated from the particle number fluctuations of spherical sub-boxes presented as functions of the inverse of the sub-box radii for methanol-water mixtures (big-box, 10000 molecules) with different methanol concentrations (mole-fraction). Solid straight lines are fitted to the linear regimes of the plots. Horizontal dashed lines denote the limiting KBIs obtained from direct integration of the corresponding RDFs.

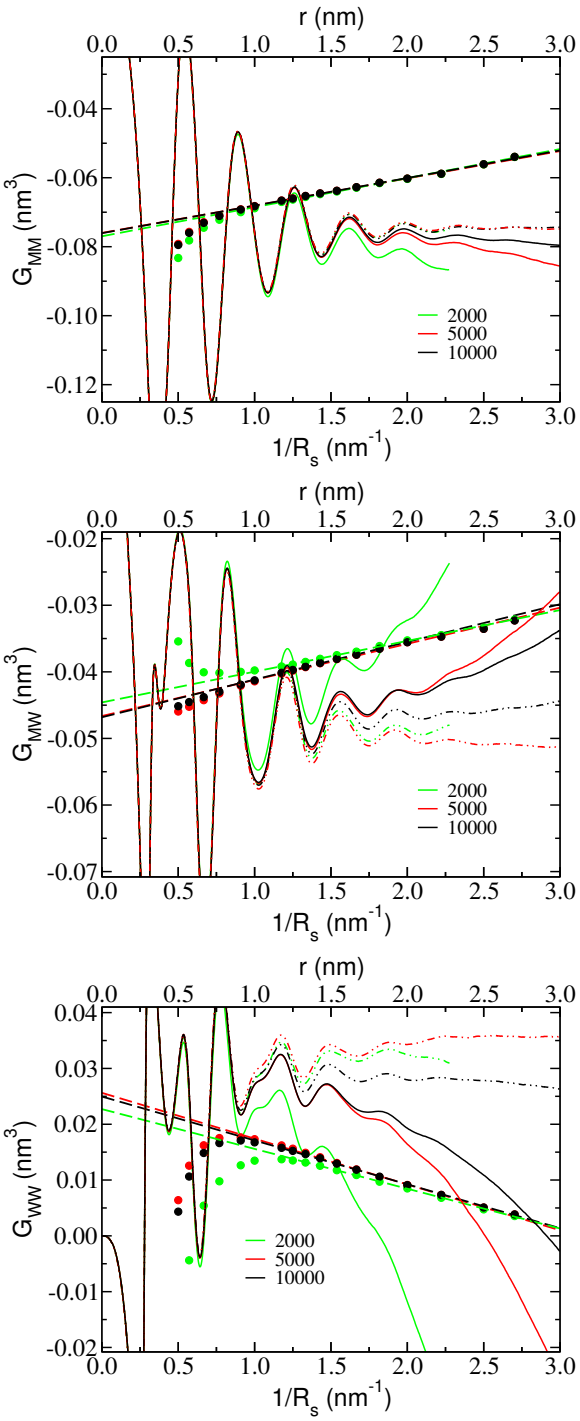


Figure 8.8: Methanol-methanol (upper panel), methanol-water (middle panel) and water-water (lower panel) SKBIs calculated from the particle number fluctuations of spherical sub-boxes presented as functions of the inverse of the sub-box radii for 50 % methanol-water mixtures with different system size (solid dots), namely with 2000, 5000 and 10000 molecules. Dashed straight lines are fitted to the linear regimes of the plots. The RKBIs are presented as solid curves (without finite-size corrections to the RDFs) and dashed curves (with finite-size corrections using Eq. 8.5). The distance scale for the RKBIs is presented on the alternative, upper horizontal axis.

The asymptotic behavior of the RDF in a finite-sized system goes to a limit, which is not one.[15] If not corrected, this error leads to RKBIs with drifting asymptotes. Empirical corrections can however be made and here we follow the idea that the observed drift happens because of the fact that excess (depletion) of particle type j around particle type i at a local-scale is compensated by depletion (excess) of particle type j at long distances as the total number of particles of type j is fixed. We therefore made a correction to the RDFs, which accounts for the correct bulk density of particle type j at a distance r from particle type i depending on the excess or depletion of particle type j within a sphere of radius r around particle type i. The corrected RDFs are given by

$$\mathrm{g}_{ij}^{correct}(r) = \mathrm{g}_{ij}(r)\frac{N_j(1-\frac{(4/3)\pi r^3}{V})}{N_j(1-\frac{(4/3)\pi r^3}{V}) - \Delta N_{ij}(r) - \delta_{ij}} \tag{8.5}$$

where N_j is the number of particles of type j in the system, V is the volume of the system and $\Delta N_{ij}(r)$ is the excess number of particles of type j within a sphere of radius r around particle type i. The RKBIs calculated from the corrected RDFs do reasonably well in terms of finding a plateau and the corresponding results for a 50% methanol mixture are shown in Figure 8.8 (see dashed black (10000 molecules), red (5000 molecules) and green (2000 molecules) curves). The differences in the KBIs obtained from big and small simulation boxes using the RKBI method are significantly reduced after applying a finite-size correction but are still larger than the differences in KBIs between box sizes obtained from the sub-box method. The data in Figure 8.8 show that the limiting behavior of the methanol-water and water-water RKBIs remains system size depend despite the tail correction. The water-water RKBIs for large r are shifted up (more positive) relative to the limiting SKBI ($L_s \to \infty$). This may reflect poor converge of the water-water RDF in this system. Since tail corrections of the RDF are empirical and may lead to different limiting RKBIs, we further tested the correction proposed by Perera and Sokolić.[14] Figure 8.9 shows the water-water RKBIs (for the system with 50% methanol) corrected using method of Perera and Sokolić (dotted lines) and Eq. 8.5 (dashes lines). Clearly, the two correction methods produce different RKBIs at distances larger than 1 nm.

The entire concentration range of methanol-water solutions has been studied using small boxes with 2000 molecules. Limiting KBIs were calculated using the sub-box method (with spherical sub-boxes) and the running-integral method. The resulting

KBIs are summarized in Table 8.2. KBIs obtained from the RKBIs were not corrected for finite-size and were obtained by averaging the RKBI in three distance ranges, between 0.8 and 1.2 nm, between 1.0 and 1.4 nm and between 1.1 and 1.5 nm. It should be pointed out that one obtains quite different KBIs (deviation of 10–25 $\%$) in case the averaging is performed using a different distance range of the same span which is a consequence of the oscillating nature and inaccuracies in the long-range parts of the RDFs for small system sizes. These deviations in the KBIs give rise to large scatter in thermodynamic quantities derived from them. On the other hand Schnell's sub-box method does not exhibit this problem as small variation of the linear-extrapolation regime within the region R_s ≈0.3 to 1.3 nm (for bigger systems) and ≈0.3 to 0.9 nm (for smaller boxes) show very small or no deviation at all (results not shown). KBIs obtained from RKBIs using a finite-size correction (Eq. 8.5) show much smaller fluctuation (5–15%) both for bigger and smaller systems (data not shown).

8.3.4.2 Effect of RDF tail corrections on RKBIs of small and large urea-water systems

Urea-water mixtures with smaller boxes (2000 water molecules) were also studied. Again KBIs were calculated using the sub-box method and the running-integral method with different averaging regions. Data obtained with the sub-box method and the running-integral method are summarized in Table 8.1, values obtained from RKBIs were not corrected for finite-size. With both methods, we observe significant variations in the urea-urea KBIs for different system sizes. It is interesting to point out that the values obtained with the running-integral method using an averaging region being 1.0 to 1.4 nm are very close to the values obtained with sub-box method for the same system size. But similarly as the methanol-water systems, if the region for averaging the RKBIs is changed, the KBIs deviate largely for the smaller systems without finite-size correction. Table 8.3 shows the KBIs obtained with tail-corrected RDFs. The differences between the resulting KBIs for different system sizes are significantly smaller in this case even compared to those obtained from the sub-box method for low urea concentrations. The fluctuations in the data obtained by varying the averaging regions of the RKBIs are 5–10% after the finite-size correction.

Urea molality	G_{ij}	Integration of RDF (corrected for finite-size)						Cubic sub-box	Spherical sub-box	
		bigbox			smallbox			bigbox	bigbox	smallbox
		$0.8-1.2$	$1.0-1.4$	$1.1-1.5$	$0.8-1.2$	$1.0-1.4$	$1.1-1.5$			
6	G_{UU}	-0.049	-0.055	-0.049	-0.049	-0.055	-0.049	-0.065	-0.064	-0.082
	G_{UW}	-0.077	-0.081	-0.080	-0.071	-0.076	-0.074	-0.077	-0.078	-0.073
	G_{WW}	-0.015	-0.015	-0.015	-0.015	-0.015	-0.015	-0.017	-0.017	-0.017
8	G_{UU}	-0.065	-0.069	-0.064	-0.071	-0.079	-0.075	-0.074	-0.075	-0.092
	G_{UW}	-0.077	-0.081	-0.080	-0.071	-0.076	-0.074	-0.074	-0.075	-0.069
	G_{WW}	-0.012	-0.012	-0.011	-0.013	-0.013	-0.012	-0.014	-0.014	-0.015
10	G_{UU}	-0.079	-0.086	-0.082	-0.082	-0.093	-0.091	-0.089	-0.090	-0.096
	G_{UW}	-0.067	-0.070	-0.069	-0.066	-0.070	-0.068	-0.068	-0.068	-0.065
	G_{WW}	-0.011	-0.011	-0.010	-0.011	-0.012	-0.011	-0.013	-0.013	-0.013
12	G_{UU}	-0.086	-0.094	-0.091	-0.087	-0.094	-0.090	-0.095	-0.096	-0.099
	G_{UW}	-0.062	-0.065	-0.063	-0.063	-0.068	-0.067	-0.063	-0.063	-0.062
	G_{WW}	-0.010	-0.010	-0.009	-0.010	-0.009	-0.008	-0.012	-0.012	-0.011

Table 8.3: Limiting KBIs for urea-water mixtures (in units nm^3) obtained from direct integration of RDFs (Eq. 8.2) after the correction for the finite-size (Eq. 8.5) applied to the corresponding RDFs and from particle number fluctuations (Eq. 8.4) using cubic sub-boxes and spherical sub-boxes (same data as in Table 8.1) by extrapolating to the thermodynamic limit (Eq. 8.3). Data are shown for big and small system sizes. For direct integration of RDFs three different regions for averaging were chosen, namely from 0.8 to 1.2 nm (0.8–1.2), from 1.0 to 1.4 nm (1.0–1.4) and from 1.1 to 1.5 nm (1.1–1.5).

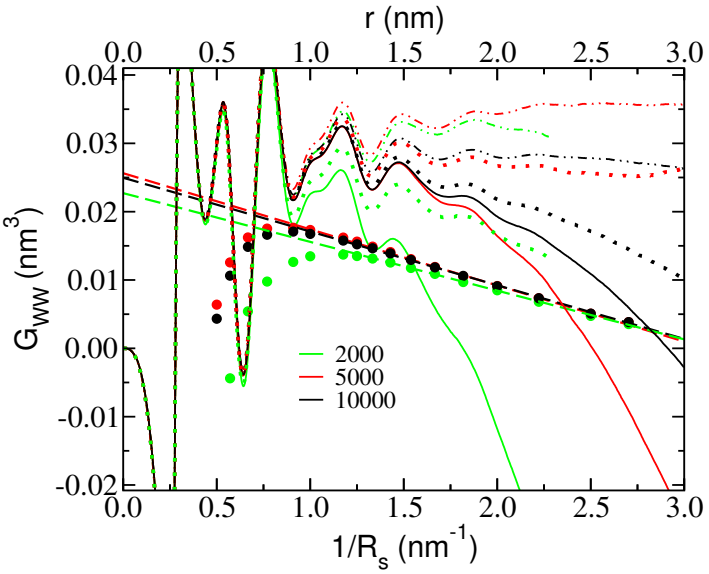


Figure 8.9: Tail corrected water-water RKBIs for 50 % methanol-water mixtures with different system sizes (2000, 5000 and 10000 molecules). The uncorrected RKBIs are presented as solid curves, RKBIs corrected with Eq. 8.5 as dashed curves and RKBIs corrected with the method of Ref.[14] as dotted curves. The distance scale for the RKBIs is presented on the upper horizontal axis. The SKBIs (thick dots) and their linear fits are also shown.

8.4 Conclusions

Kirkwood-Buff integrals (KBIs) of aqueous solution mixtures provide a link between local solution structure and global thermodynamic solution properties. In recent years, KBIs have been used in force field parameterization [4–12] as well as in simulation studies aiming to relate ion-specific thermodynamic changes and atomic scale correlations. [16,35,36] Notwithstanding the powerful scope of Kirkwood-Buff theory, sampling of the particle number fluctuations that determine the KBIs requires to address convergence issues and finite-size effects. In this paper, we have performed molecular dynamics simulations of methanol-water and urea-water mixtures with earlier-published Kirkwood-Buff-derived force fields. KBIs have been calculated by integrating the radial distribution functions over the volume (running-KBI method) as well as by direct analysis of particle number fluctuations in small sub-boxes embedded in the overall NpT simulation box (sub-box method). While in the former approach the thermodynamic limiting value of the KBI is obtained from the running integral in the limit of large distances, it is obtained in the latter approach by examining different sub-box sizes and applying an analytical finite-size scaling relation. [19] We find that with either of these two approaches converged KBIs can only be obtained from molecular dynamics simulations on sufficient long time scales, typically of the order of 100 nanoseconds. Previous simulations of aqueous solutions have all used significantly shorter simulations (several nanoseconds) to obtain the KBIs. A possible explanation of the slow time convergence may be sought in the micro-heterogeneous nature of aqueous solutions, which leads to slow domain-like dynamics. We furthermore find that the running-KBI method suffers from stronger finite-size artifacts than the sub-box method. Simulations with different systems sizes (2000, 5000, 10000 molecules) indicate that the sub-box method provides limiting KBIs, which are in good agreement for all three system sizes. The running-KBI method however provides estimates of the limiting KBIs, which, depending on system size and the region for averaging the running-KBIs, show larger variations which are more severe for smaller systems. If however a finite-size correction (Eq. 8.5) for the RDFs is carried out, the running-integral method provides improved KBIs with smaller variations upon varying the region for averaging the running-KBIs. For larger systems and for longer simulation runs both the methods seem to produce very similar results with the sub-box method being more precise in most cases.

We finally point out that the running-KBI method provides a very powerful tool to analyze contributions of local correlations to the thermodynamic quantities. Examples of these include microscopic explanations for ion-specific osmotic properties of aqueous electrolyte solutions[16,35] and salting-in and salting-out mechanisms of water soluble polymers and small peptides[36,37] which cannot be provided by the sub-box method.

8.5 Acknowledgements

The authors wish to thank Debashish Mukherji for useful discussions during the course of this work. This research was supported by the German Research Foundation (DFG) within the Cluster of Excellence 259 "Smart Interfaces – Understanding and Designing Fluid Boundaries".

Bibliography

[1] Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

[2] Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.

[3] Ben-Naim, A. *J. Chem. Phys.* **1977**, *67*, 4884.

[4] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *118*, 10663.

[5] Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003**, *107*, 3891.

[6] Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2004**, *121*, 2180.

[7] Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005**, *109*, 15080.

[8] Lee, M. E.; van der Vegt, N. F. A. *J. Chem. Phys.* **2005**, *122*, 114509.

[9] Gee, M. B.; Cox, N. R.; Jiao, Y. F.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. *J. Chem. Theory Comp.* **2011**, *7*, 1369.

[10] Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006**, *27*, 1477.

[11] Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N. F. A. *J. Chem. Theory Comput.* **2012**, *8*, 1802.

[12] Fyta, M. ; Netz, R. R. *J. Chem. Phys.* **2012**, *136*, 124103.

[13] Lebowitz, J. L.; Percus, J. K. *Phys. Rev.* **1961**, *124*, 1673.

[14] Perera, A.; Sokolić, F. *J. Chem. Phys.* **2004**, *121*, 11272.

[15] Perera, A.; Zoranić, L.; Sokolić, F.; Mazighi, R. *J. Mol. Liq.* **2011**, *159*, 52.

[16] Hess, B.; van der Vegt, N. F. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13296.

[17] Christensen, S.; Peters, G. H.; Hansen, F. Y.; O'Connell, J. P.; Abildskov, J. *Mol. Simul.* **2007**, *33*, 449.

[18] Wedberg, R.; Peters, G. H.; Abildskov, J. *Fluid Phase Equilib.* **2008**, *273*, 1.

[19] Schnell, S. K.; Liu, X.; Simon, J.-M.; Bardow, A.; Bedeaux, D.; Vlugt, T. J. H.; Kjelstrup, S. *J. Phys. Chem. B* **2011**, *115*, 10911.

[20] Schnell, S. K.; Vlugt, T. J. H.; Simon, J.-M.; Bedeaux, D.; Kjelstrup, S. *Chem. Phys. Lett.* **2011**, *504*, 199.

[21] Hill, T. L. *Thermodynamics of Small Systems, Part 1*; Benjamin: New York, 1963.

[22] Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, *7*, 306.

[23] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

[24] Nose, S. *Mol. Phys.* **1984**, *52*, 255.

[25] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

[26] Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.

[27] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

[28] Bowron, D. T.; Finney, J. L.; Soper, A. K. *J. Phys. Chem.* **1998**, *102*, 3551.

[29] Dixit, S.; Crain, J.; Poon, W. C.; Finney, J. L.; Soper, A. K. *Nature* **2002**, *416*, 829.

[30] Allison, S. K.; Fox, J. P.; Hargreaves, R.; Bates, S. P. *Phys. Rev. B* **2005**, *71*, 024201.

[31] Perera, A.; Sokolić, F.; Almasy, L.; Koga, Y. *J. Chem. Phys.* **2006**, *124*, 124515.

[32] Sokolić, F.; Idrissi, A.; Perera, A. *J. Chem. Phys.* **2002**, *116*, 1636.

[33] Stumpe, M. C.; Grubmüller, H. *J. Phys. Chem. B* **2007**, *111*, 6220.

[34] Dougan, L.; Bates, S. P.; Hargreaves, R.; Fox, J. P.; Crain, J.; Finney, J. L.; Réat, V.; Soper, A. K. *J. Chem. Phys.* **2004**, *121*, 6456.

[35] Ganguly, P.; Schravendijk, P.; Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2011**, *115*, 3734.

[36] Algaer, E.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2011** *115*, 13781.

[37] Mukherji, D.; van der Vegt, N. F. A.; Kremer, K. *J. Chem. Theory Comp.* **2012** *8*, 3536.

9 Conclusions and Outlook

This thesis has served to explain the potential applications of the Kirkwood-Buff theory of solutions to develop simplified models for computer simulations of aqueous solutions and also to unravel the physical mechanisms behind the cosolvent effects on the solubility of the solutes in water mixtures. Kirkwood-Buff (KB) theory relates the particle number fluctuations at local scale in a solution to the global thermodynamic properties of the solutions in terms of the compressibility, partial volumes or the variation in the solvation free energy of the solutes. This thesis mainly focuses on three interrelated topics: a) parametrization of structurally and thermodynamically consistent coarse-grained models using Kirkwood-Buff theory, b) applications of the KB theory to understand the ion-pairing mechanism in biologically relevant salt solutions, and c) technical issues in the application of the KB theory in computer simulations and possible fixes.

Single-site coarse-grain models for urea-water and benzene-water solutions have been developed by using the structure based Iterative Boltzmann Inversion (IBI) method for coarse-graining combined with the KB theory. The models, namely KB-IBI coarse-grained force-fields, preserve the local pair-correlations of the solutions and also show correct variation in the chemical potential or activity of the solutes (urea or benzene) with varying solute concentrations. The KB-IBI models for the urea-water mixtures have further been applied to study the salting-in of benzene in urea-water solutions where benzene-water and benzene-urea interactions have been parameterized using KB-IBI method. The limitation in the transferability of the KB-IBI force-fields for modeling of the preferential interactions in ternary solutions has been tested rigorously. It has been found that the KB-IBI models work better than the IBI models to reproduce the thermodynamic quantities related to solvation but the models remain transferable only when the fluctuation in the solution concentration is not too high. The representability of the structure-based coarse-graining models has been discussed in terms of the system pressure, compressibility and the variation in the solvation free-energy of the solution components for the systems of pure water, binary urea-water mixtures and ternary benzene-urea-water mixtures. It has been shown that KB-IBI method combined with a correction for the systems pres-

sure (as coarse-grained systems produce very high pressure in general) can serve to find an optimal compromise between these quantities and also KB-IBI method leads to faster convergence of these quantities including the potential energy of the systems. Like other structure-based coarse-graining force-fields KB-IBI models also show very strong state-point dependence as the potentials account for the many-body correlations in a solution. But KB-IBI models have been found to be very effective in modeling single-phase aqueous solutions with thermodynamic consistency. As KB-IBI models accurately preserve the preferential solvation between the solution components, the KB-IBI method can be thought as the first-step towards modeling the cosolvent-driven conformational changes of the macromolecules in aqueous co-solvent solutions. Temperature- and chemical transferability of the KB-IBI models are yet to be tested.

KB theory has been used to study the ion-pairing mechanism between dimethyl phosphate and Hofmeister alkali cations using all-atom simulations. The simulation results have been compared directly with the experimental results quantitatively and it has been found that a water-mediated ion-pairing mechanism between the cations and the anion prevails over the direct-pairing mechanism and determines the difference in the thermodynamic activity of the different salts. On the contrary, for chloride salts, direct-pairing between the cations and the chloride ion plays a more significant role. The effect of the alkali chloride salts on the solubility of benzene in water has also been studied with all-atom simulations. Quantitative studies from the simulations related to the experimentally observables lead to the inference that the direct correlations between the cations and benzene determine the difference in the solubility of benzene in different salt solutions. The pairing between the ions or between the ions and the water molecules does not play any significant role. Benzene-water and water-water correlations do not show any ion-specific changes for different salt solutions. The application of the KB theory to understand molecular liquids can further be extended to understand the salt-induced conformational changes, for example, protein denaturation by Hofmeister salts can be studied by the means of the preferential solvation between the solute residues and the salt-ions. Effects of the osmolytes on the hydrophobic-hydrophilic interactions in aqueous solutions can also be analyzed using KB theory. But the calculation of the Kirkwood-Buff integrals from the computer simulations, which are the key-quantities in the KB theory, does come with a number of technical issues related to the system sizes and the convergence of the local particle-fluctuations with time. There have been many attempts to fix the

tail of the pair-correlation functions to account for the effect of the finite-size of the system on the Kirkwood-Buff integrals. But the results strongly depend on the choice of these empirical fixes. Alternatively, more precise Kirkwood-Buff integrals can be calculated from the particle fluctuations in small systems embedded in a larger system but the time convergence of the KBIs calculated by this method does not find any significant improvement. A more detailed understanding about the relaxation of the correlated movements of the molecules over time in the aqueous cosolvent solutions and a robust theoretical analysis for correcting the Kirkwood-Buff integrals would attract much more accurate analysis of the molecular mechanisms of the solvation thermodynamics in the aqueous solutions with the help of the Kirkwood-Buff theory.

Curriculum Vitae

Name: Pritam Ganguly

Place of Birth: Kadamtala, Howrah, India

Date of Birth: 17.11.1987

Nationality: Indian

Education:

1997 – 2005: High School: West Bengal Council of Higher Secondary Education, India.

2005 – 2008: Bachelor of Science (Honours in Physics), University of Calcutta, India.

2008 – 2010: Master of Science (Physics), Indian Institute of Technology- Guwahati, India.

2010 – 2013: Center of Smart Interfaces, Technische Universität Darmstadt, Germany.

_____

(Pritam Ganguly)

Pritam Ganguly
Center of Smart Interfaces
Alarich-Weiss-Straße 10
64287 Darmstadt

Tel.: 06151 16 4478
E-mail: ganguly@csi.tu-darmstadt.de

Darmstadt, den 30. Oktober 2013

Erklärung

Ich erkläre hiermit, noch keinen Promotionsversuch unternommen zu haben.

_____

(Pritam Ganguly)

Pritam Ganguly
Center of Smart Interfaces
Alarich-Weiss-Straße 10
64287 Darmstadt

Tel.: 06151 16 4478
E-mail: ganguly@csi.tu-darmstadt.de

Darmstadt, den 30. Oktober 2013

Eidesstattliche Erklärung

Ich erkläre mich hiermit an Eides Statt, dass ich meine Dissertation selbständig und nur mit den angegebenen Hilfen angefertigt habe.

_____

(Pritam Ganguly)